```
OM protein - protein search, using sw model

          GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:        May 18, 2004, 11:21:33 ; Search time 24.9726 Seconds
                                (without alignments)
                                5623.200 Million cell updates/sec

Title:          US-08-353-485-2
Perfect score:  2641
Sequence:       1 MRKLNSLFSLAVLLSLLCWG.........QNLTGSAVGQXVTLKWDAPN 497
Scoring table:  BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched:       1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
             1:_geneseqp1980s:*
             2:_geneseqp1990s:*
             3:_geneseqp2000s:*
             4:_geneseqp2001s:*
             5:_geneseqp2002s:*
             6:_geneseqp2002as:*
             7:_geneseqp2003bs:*
             8:_geneseqp2004s:*

                        SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result          %      Query
No.    Score  Match  Match  Length  DB  ID        Description
----------------------------------------------------------------------
 1     2641  100.0         497    2  AAR96025  Aar96025 P. gingiv
 2     2641  100.0         497    2  AAW69483  Aaw69483 Haemagglu
 3     2641  100.0        2628    2  AAR96030  Aar96030 P. gingiv
 4     2641  100.0        2628    2  AAW69488  Aaw69488 Haemagglu
 5      951.5  36.0       1706    2  AAW24786  Aaw24786 PrtR anti
 6      944.5  35.8       1704    2  AAR70188  Aar70188 Arg-gingi
 7      944.5  35.8       1704    2  AAW34843  Aaw34843 Arg-gingi
 8      944.5  35.8       1704    4  AAY67396  Aay67396 Arg-gingi
 9      944.5  35.8       1704    4  AAU08938  Aau08938 P. gingiv
10      943.5  35.7       1687    2  AAR96033  Aar96033 P. gingiv
11      943.5  35.7       1687    2  AAW69495  Aaw69495 Haemagglu
12      816    30.9       1358    2  AAR96032  Aar96032 P. gingiv
13      816    30.9       1358    2  AAW69494  Aaw69494 Haemagglu
14      812.5  30.8       1732    2  AAR96029  Aar96029 P. gingiv
15      812.5  30.8       1732    2  AAW24787  Aaw24787 PrtK anti
16      812.5  30.8       1732    2  AAW69487  Aaw69487 Haemagglu
17      806.5  30.5       1087    2  AAR96028  Aar96028 P. gingiv
18      806.5  30.5       1087    2  AAW69486  Aaw69486 Haemagglu
19      791.5  30.0        419    4  AAU03572  Aau03572 P. gingiv
20      662.5  25.1        419    4  AAU03574  Aau03574 P. gingiv
21      546.5  20.7        450    2  AAR96021  Aar96021 P. gingiv
22      546.5  20.7        450    2  AAW69489  Aaw69489 Haemagglu
23      546.5  20.7        456    2  AAR96023  Aar96023 P. gingiv
24      546.5  20.7        456    2  AAR96022  Aar96022 P. gingiv
25      546.5  20.7        456    2  AAW69491  Aaw69491 Haemagglu
26      546.5  20.7        456    2  AAW69490  Aaw69490 Haemagglu
27      522    19.8        970    2  AAR72458  Aar72458 Porphyrom
28      495.5  18.8        991    2  AAR77313  Aar77313 Porphyrom
29      458.5  17.4        439    2  AAR96024  Aar96024 P. gingiv
30      458.5  17.4        439    2  AAW69492  Aaw69492 Haemagglu
31      457.5  17.3        231    4  AAU03575  Aau03575 P. gingiv
32      425.5  16.1        377    2  AAY34359  Aay34359 Porphyrom
33      379    14.4        312    2  AAY34484  Aay34484 Porphyrom
34      337    12.8        921    2  AAY34522  Aay34522 Porphyrom
35      337    12.8        922    2  AAY34521  Aay34521 Porphyrom
36      337    12.8        925    2  AAY34520  Aay34520 Porphyrom
37      337    12.8        938    2  AAY34392  Aay34392 Porphyrom
38      272.5  10.3        148    3  AAB14942  Aab14942 Porphyrom
39      213.5   8.1        293    2  AAY34483  Aay34483 Porphyrom
40      213.5   8.1        299    2  AAY34358  Aay34358 Porphyrom
41      188.5   7.1        196    4  AAU03573  Aau03573 P. gingiv
42      177.5   6.7        135    6  ABP55081  Abp55081 Porphyrom
43      176.5   6.7        134    4  AAB49217  Aab49217 Peptide u
44      175     6.6         49    2  AAW34805  Aaw34805 Arg-speci
45      153     5.8       5291    7  ADC01014  Adc01014 Enterohae

                        ALIGNMENTS

RESULT 1
AAR96025
ID   AAR96025 standard; protein; 497 AA.
XX
AC   AAR96025;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;  95WO-US016108.
XX
PR   09-DEC-1994;  94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30649.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 47-51; 153pp; English.
XX
CC   A portion (AAR96025) of P. gingivalis 381 haemagglutinin hagA (see also
CC   AAR96030) was identified as the product of clone ST2 (AAT30649) isolated
CC   from a P. gingivalis 318 genomic library. The haemagglutinin can be obtd.
CC   from transformed host cells and used as a vaccine to protect humans or
CC   animals against periodontal disease. Expression in Salmonella cells
CC   allows prodn. of a live vaccine. The haemagglutinin can also be used to
CC   detect the presence of anti-P. gingivalis antibodies and to raise
CC   monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-2003 to
CC   standardise OS field)
XX
SQ   Sequence 497 AA;
```

Query Match   100.0%; Score 2641; DB 2; Length 497;
Best Local Similarity 100.0%; Pred. No. 2.5e-194;
Matches 497; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM  60

Qy   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120

Qy  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIIYVGEGVSKGNDYVVEAGKTYHFT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIIYVGEGVSKGNDYVVEAGKTYHFT 180

Qy  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240

Qy  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300

Qy  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360

Qy  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGGSSNDYTYTV 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGGSSNDYTYTV 420

Qy  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480

Qy  481 TGSAVGQKVTLKWDAPN 497
        |||||||||||||||||
Db  481 TGSAVGQKVTLKWDAPN 497
```

RESULT 2
AAW69483
ID AAW69483 standard; protein; 497 AA.
XX
AC AAW69483;
XX
DT 22-DEC-1998 (first entry)
XX
DE Haemagglutinin protein hagA.
XX
KW Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS Porphyromonas gingivalis.
XX
PN US5824791-A.
XX
PD 20-OCT-1998.
XX
PF 11-DEC-1995; 95US-00570311.
XX
PR 08-SEP-1988; 8US-00241640.
PR 25-JAN-1991; 91US-00647119.
PR 09-DEC-1994; 94US-00353485.
XX
PA (UYFL) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR WPI; 1998-582627/49.
DR N-PSDB; AAV58870.
XX

PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT protease poly:peptide(s)).
XX
PS Claim 1; Col 37-44; 101pp; English.
XX
CC This sequence is encoded by a Porphyromonas gingivalis gene of the
CC invention. This sequence represents the hagA haemagglutinin protein. The
CC polypeptides are used to produce antibodies to organisms associated with
CC periodontal disease. The antibodies are also used in purification and
CC identification procedures. The genes and polypeptides are used as
CC vaccines against periodontal disease
XX
SQ Sequence 497 AA;

Query Match   100.0%; Score 2641; DB 2; Length 497;
Best Local Similarity 100.0%; Pred. No. 2.5e-194;
Matches 497; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM  60

Qy   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120

Qy  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIIYVGEGVSKGNDYVVEAGKTYHFT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIIYVGEGVSKGNDYVVEAGKTYHFT 180

Qy  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240

Qy  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300

Qy  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360

Qy  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGGSSNDYTYTV 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGGSSNDYTYTV 420

Qy  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480

Qy  481 TGSAVGQKVTLKWDAPN 497
        |||||||||||||||||
Db  481 TGSAVGQKVTLKWDAPN 497
```

RESULT 3
AAR96030
ID AAR96030 standard; protein; 2628 AA.
XX
AC AAR96030;
XX
DT 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)
XX
DE P. gingivalis haemagglutinin hagA.
XX
KW Haemagglutinin; hagA; periodontal disease; vaccine; antibody.
XX
OS Porphyromonas gingivalis; strain 381.
XX
FH Key   Location/Qualifiers
FT Peptide   5..21
FT   /label= Sig_peptide

```
                 XX
                 PN   WO9617936-A2.
                 XX
                 PD   13-JUN-1996.
                 XX
                 PF   11-DEC-1995;   95WO-US016108.
                 XX
                 PR   09-DEC-1994;   94US-00353485.
                 XX
                 PA   (UYFL ) UNIV FLORIDA.
                 PA   (UABR-) UAB RES FOUND.
                 XX
                 PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
                 XX
                 DR   WPI; 1996-287181/29.
                 DR   N-PSDB; AAT30654.
                 XX
                 PT   Porphyromonas gingivalis genes and proteins - used in the detection and
                 PT   vaccination against periodontal disease.
                 XX
                 PS   Claim 6; Page 93-101; 153pp; English.
                 XX
                 CC   P. gingivalis 381 haemagglutinin hagA (AAR96030) was identified as the
                 CC   product of the hagA gene (AAT30654) isolated as an EcoRV fragment of
                 CC   genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
                 CC   24), can be obtd. from transformed host cells and used as a vaccine to
                 CC   protect humans or animals against periodontal disease. Expression in
                 CC   Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
                 CC   also be used to detect the presence of anti-P. gingivalis antibodies and
                 CC   to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-
                 CC   2003 to standardise OS field)
                 XX
                 SQ   Sequence 2628 AA;

Query Match             100.0%;  Score 2641;  DB 2;  Length 2628;
Best Local Similarity   100.0%;  Pred. No. 2.9e-193;
Matches  497;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy           1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60

Qy          61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120

Qy         121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYYVEAGKTYHFT  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYYVEAGKTYHFT  180

Qy         181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240

Qy         241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300

Qy         301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDKPAPMNLVK  360

Qy         361  SEGVKLPAPYQERTIDLSAYAGQQYYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361  SEGVKLPAPYQERTIDLSAYAGQQYYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420

Qy         421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480

Qy         481  TGSAVGQKVTLKWDAPN  497
                |||||||||||||||||
Db         481  TGSAVGQKVTLKWDAPN  497
```

```
RESULT 4
AAW69488
ID   AAW69488 standard; protein; 2628 AA.
XX
AC   AAW69488;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagA.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58875.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 91-110; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 2628 AA;

Query Match             100.0%;  Score 2641;  DB 2;  Length 2628;
Best Local Similarity   100.0%;  Pred. No. 2.9e-193;
Matches  497;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy           1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60

Qy          61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120

Qy         121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYYVEAGKTYHFT  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYYVEAGKTYHFT  180

Qy         181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240

Qy         241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300

Qy         301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDKPAPMNLVK  360
```

```
Db   301  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360

Qy   361  SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420

Qy   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480

Qy   481  TGSAVGQKVTLKWDAPN 497
          |||||||||||||||||
Db   481  TGSAVGQKVTLKWDAPN 497
```

RESULT 5
AAW24786
ID   AAW24786 standard; protein; 1706 AA.
XX
AC   AAW24786;
XX
DT   17-OCT-2003 (revised)
DT   25-NOV-1997 (first entry)
XX
DE   PrtR antigenic protein complex.
XX
KW   Periodontal disease; cell surface protein; thiol protease; endopeptidase;
KW   PrtR; PrtR45; PrtR44; PrtR15; PrtR17; PrtR27; haemagglutinin; adhesin;
KW   therapy; diagnosis; vaccine; antigen.
XX
OS   Porphyromonas gingivalis; strain W50.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..227
FT   /label= Pro-pro_peptide
FT   Cleavage-site   227..228
FT   Protein   228..719
FT   /label= PrtR45
FT   /note= "45 kDa Arg-specific thiol protease"
FT   Cleavage-site   719..720
FT   Protein   720..1138
FT   /label= PrtR44
FT   /note= "44 kDa adhesin"
FT   Cleavage-site   1138..1139
FT   Protein   1139..1273
FT   /label= PrtR15
FT   /note= "15 kDa adhesin"
FT   Cleavage-site   1273..1274
FT   Protein   1274..1431
FT   /label= PrtR17
FT   /note= "17 kDa adhesin"
FT   Cleavage-site   1431..1432
FT   Protein   1432..1706
FT   /label= PrtR27
FT   /note= "27 kDa adhesin"
XX
PN   WO9716542-A1.
XX
PD   09-MAY-1997.
XX
PF   30-OCT-1996; 96WO-AU000673.
XX
PR   30-OCT-1995; 95AU-00006275.
XX
PA   (UYME) UNIV MELBOURNE.
PA   (VICT-) VICTORIAN DAIRY IND AUTHORITY.
XX
PI   Reynolds EC, Bhogal PS, Slakeski N;
XX
DR   WPI; 1997-272112/24.
DR   N-PSDB; AAT78850.
XX

PT   New antigenic protein complex from Porphyromonas gingivalis - comprising
PT   Arg- and Lys- specific thiol endo-peptidase(s), used in the detection,
PT   prevention and treatment of periodontal disease.
XX
PS   Example 1; Fig 8b; 68pp; English.
XX
CC   A PrtR-PrtK cell surface protein of Porphyromonas ginivalis (PG)
CC   comprises a 300 kDa complex composed a 45 kDa arginine-specific thiol
CC   protease and 44, 15, 17 and 27 kDa adhesins encoded by the prtR gene
CC   (AAT78850), and a 148 kDa lysine-specific thiol protease and 39, 15 and
CC   44 kDa adhesins (see AAW24787) encoded by the prtK gene (AAT78851). A
CC   claimed antigenic complex comprises at least one multimeric protein
CC   complex of PrtR and PrtK each containing at least one adhesin domain, the
CC   complex having a mol. wt. of over 200 kDa, and preferably comprises all 9
CC   proteins of the PrtR-PrtK complex (see also AAW24780-85). It can be used
CC   in a claimed composition to elicit an immune response directed against
CC   PG, and in a claimed method of reducing the prospect of PG infection
CC   and/or severity of disease. Antibodies directed against the complex are
CC   claimed for use in treating PG infection. Unlike whole PG cells or other
CC   previously prepared antigens based on fimbriae or the capsule, the PrtR-
CC   PrtK complex or component parts are safe and effective antigens. (Updated
CC   on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1706 AA;

Query Match   36.0%;   Score 951.5;   DB 2;   Length 1706;
Best Local Similarity 31.9%;   Pred. No. 1.3e-63;
Matches 237;   Conservative 74;   Mismatches 156;   Indels 0;   Gaps 20;

```
Qy    22  TAAAQG------GPKTAPSVTHQAVQK--GIR--TSKVKDLRDPIPAGMARIILEAH   68
          ||||||        |  :  :::::||    ||    ::  | : :||  |||  ||||
Db   671  TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy    69  DYWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
          | | :|:|| :|:|||| || ::| ::      |   ||  | |  ||:||:|||  :||
Db   731  DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy   127  TNFVLDGTASADIPAGTYDYVIINPNGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG  185
          ||::::|||| :|||||||: :|  :    :|: | :    ||| ||||||||||   |
Db   790  TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy   186  PGDAASVV------                                               193
          .|| :  :
Db   850  SGDGTELTISEGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS  909

Qy   194  ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTYVLNE    235
                ||| |:||||||| || | | :| |:| ||:||:           :||| |:|
Db   910  PKVCKDVTVEGSNEFAPVQNLTGSAVGKVTLKWDAPNGTPNPNPNPNPNPGTTTLSE   969

Qy   236  SPDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
          | : : ||||:: |||||||| | :   |:  | | |    :| ||:||   : :|:|||
Db   970  SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1025

Qy   296  YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETL--GSD   351
          ||:||  ::|:||||: |||  |  ::: |||| ::|||| ::||| ||||||    |:
Db  1026  YLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1085

Qy   352  KPAPM------                                                  356
          .| .
Db  1086  SPEAMRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1145

Qy   357  ------NLVKS                                                  361
                 ::|:
Db  1146  ESSTHGEAPAEWTTIDADGDGDGQWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALNPDNYL 1205

Qy   362  ------                                                       361
Db  1206  ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGL   1265

Qy   362  ------EGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDV--AVSGEGSSND 415
                 | :||||||||||:||| |||| ::||||||:| :  : : ||||:  ||:|| ||||
Db  1266  STEADGAKPQSVWIERTVDLP-AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTFTD   1322
```

```
Qy   416  YTYTYVRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA  475
          ||||||   : :| |||  :||::| |||||||:| |||||||||||:|||
Db  1323  YTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN 1381

Qy   476  HVQNLTGSAVGQKVTLKWDAPN  497
          ||:|| |||||||:|||:|
Db  1382  PVKNLKAQPDGGDVVLKWEAPS 1403
```

RESULT 6
AAR70188
ID   AAR70188 standard; protein; 1704 AA.
XX
AC   AAR70188;
XX
DT   25-MAR-2003 (revised)
DT   21-SEP-1995 (first entry)
XX
DE   Arg-gingipain-2 prepolyprotein.
XX
KW   Arg-gingipain-2; gingivalis; periodontal disease; vaccine;
KW   arginine-specific protease.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key   Location/Qualifiers
FT   Protein   228..719
FT   /note= "Protease"
FT   /note= "corresponds to Arg-gingipain-1"
FT   Region   720..1091
FT   /label= Hemagglutinin
FT   Region   1092..1429
FT   /label= Hemagglutinin
FT   Region   1430..1704
FT   /label= Hemagglutinin
XX
PN   WO9507286-A1.
XX
PD   16-MAR-1995.
XX
PF   09-SEP-1994; 94WO-US010283.
XX
PR   10-SEP-1993; 93US-00119361;
PR   21-OCT-1993; 93US-00141324;
PR   24-JUN-1994; 94US-00265441;
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J, Potempa J, Barr PJ, Pavloff N;
XX
DR   WPI; 1995-123373/16.
DR   N-PSDB; AAQ83489.
XX
PT   DNA encoding Arg-gingipain proteins - used to develop prods. for
PT   detection, treatment and prevention of periodontal disease.
XX
PS   Disclosure; Page 70-77; 89pp; English.
XX
CC   A low mol.wt. arginine-specific gingipain (AG-1) and high mol.wt. AG (AG-
CC   2) were isolated from P. gingivalis strains H66 (ATCC 33277) and W50
CC   (ATCC 53973). The sequences of the proteins were used to design PCR
CC   primers and probes to isolate AG DNA. Lambda DASH and lambda ZAP
CC   libraries were screened with a probe based on amino acids 11-22 of the AG
CC   protein to obtain DNA encoding AG-1 (AAQ83484) and AG-2 (AAQ83489). AG-2
CC   is a prepolyprotein incorporating AG-1. (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 1704 AA;

Query Match   35.8%; Score 944.5; DB 2; Length 1704;
Best Local Similarity   31.9%;
Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

```
Qy    22  TAAAQG-------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH   68
          |::||      ::: ::  | | :       |||      |||
Db   671  TATTQQQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH  730

Qy    69  DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP  126
          |||:||:||: :|||||:||| :|::   | :  | : ||  |::|||:|| ||| :|||
Db   731  DVWNDGSGYQIILDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP  789

Qy   127  TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG  185
          ||:::||||| :||||||| : :   ::  :|:|| |    ||||:||| ||||  ::||
Db   790  TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG  849

Qy   186  PGDAASVV------                                               193
          ||::
Db   850  SGDGTELTISEGGGSDYTYTVVRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS  909

Qy   194  ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------DKRTYVLNESF      237
                 :||||||||||||| :::  :||:||| ||||:      || ||:|:|||
Db   910  PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSSSF    969

Qy   238  DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL  297
          : :  ||| : |||||||| | |   | | :: |   || |  | |:||  :| ||||||
Db   970  E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy   298  VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS------  350
          :|| : :|:||||:  | :   ::  ||| |  ::|||: ||||  ||||| :
Db  1026  ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy   351  ------DKPAPMNLV------KSEGVKL------                           366
                ::       ||:
Db  1086  EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETPES 1145

Qy   367  --PAPYQERTID-----------LSAIYAGQQV-                           385
            :|: | : ||                :||:|||
Db  1146  STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy   386  ------                                                       385
Db  1206  KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGLST  1265

Qy   386  ------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV                  420
                ||| |||| ||  |::||:|:   ||| |: || ||||
Db  1266  EANGAKFQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYYTYTV 1325

Qy   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
          |||   : ::|  |||::: |||| :: ||||||||||||| |:|| || : :
Db  1326  YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1384

Qy   481  TGSAVGQKVTLKWDAPN                                            497
          :: :|| :|||||:||
Db  1385  KAQPDGGDVVLKWEAPS                                           1401
```

RESULT 7
AAW34843
ID   AAW34843 standard; protein; 1704 AA.
XX
AC   AAW34843;
XX
DT   03-JUN-1998 (first entry)
XX
DE   Arg-gingipain high molecular weight prepolyprotein sequence.
XX
KW   Arg-specific gingipain protease; gingivalis; periodontal disease;
KW   vaccine; infection.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key   Location/Qualifiers
FT   Protein   1..227
FT   /note= "precursor protein"
XX
PN   WO9734629-A1.

```
XX
PD   25-SEP-1997.
XX
PF   21-MAR-1997;   97WO-US004635.
XX
PR   22-MAR-1996;   96US-0013945P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
PA   (MORE-) MOREHOUSE SCHOOL MEDICINE.
XX
PI   Potempa J, Travis J, Genco C;
XX
DR   WPI; 1997-479993/44.
DR   N-PSDB; AAT93872.
XX
PT   Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
PT   useful for protecting animals and humans from gingivalis and periodontal
PT   diseases.
XX
PS   Disclosure; Page 68-73; 95pp; English.
XX
CC   The present sequence represents an arginine-specific protease of
CC   Porphyromonas gingivalis. The following peptides, derived from Arg- and
CC   Lys-specific high molecular weight proteases, offer protection against
CC   infection: YTYTVYRDGK IKEGLTATTE DDGVATGNHE YCVEKYTAGS VSPKVC (I);
CC   YTPVEEKQNG RMIVIVAKKY (II); QLPFIFDVAC VNGDFLFSMP CFAEALMRAQ (III);
CC   GEPNEYQPVS NLTATTQGQK VTLKWDAPSTK (IV); GNHEYCEVKYTAGVSPKVC KDVTV (V);
CC   RMFMNYEPGR YTPVEEKQNG (VI); TFAGFEDTYK RMFMNYEPGR (VII); DYTYTVYRDG
CC   TKIKEGLTAT TFEEDGVATG NMEYCVCVKY TAGVSPKVC (VIII); YTYTVYRDGT KIKEGLTATTF
CC   EEDG (IX); RDGTKIKEGL TATTFEEDGV ATGN (X); KIKEGLTATT FEEDGVATGN HEY (XI)
CC   ; KWDAPNGTPN PNPN PNPNPN PGTTLSE (XII); and YTPVEEKENG RMIVIVAKKY
CC   (XIII). They are used in vaccines to protect animals, including humans,
CC   from gingivitis and/or periodontal diseases
XX
SQ   Sequence 1704 AA;

Query Match             35.8%;  Score 944.5;  DB 2;  Length 1704;
Best Local Similarity   31.9%;  Pred. No. 4.4e-63;
Matches  235;  Conservative  75;  Mismatches 160;  Indels 267;  Gaps  20;

Qy      22 TAAAQG--------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH  68
                         ||||||:| :    :  ||       |  |||| :   :: |||||
Db     671 TATTQGGKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy      69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
            ||::|:|||:|:|||: :        |:   |||:|||||||:||||||||||||:|:||
Db     731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy     127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKIYHFTVQRQG 185
            |||:| |||||||||||||||||:::|:| |||| :||||||||:||||:|||:|::
Db     790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYTHFLMKKMG 849

Qy     186 PGDAASVV--------------------------------------------------- 193
            |:|||:||
Db     850 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909

Qy     194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------------DKRTYVLNESF 237
                  ||||||||||||:||||||||||||:||||||            ||:|||||||
Db     910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSESF 969

Qy     238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
            ||||||||||||||||:||||||||||||||||||||||||||:||::||:|:| | |:
Db     970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy     298 VTPKVTVPENGKLSYWSSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETLGS-------- 350
            ||:| ::||| |||:|| |  |   |:|  ||| : ||  |||:||  |:|
Db    1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy     351 ----------DKPAPMNLV-----------------------KSEGVKL---------- 366
                      |||||::||                       |: :
Db    1086 EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145

Qy     367 ---PAPYQERTID-----------------------LSAYAGQQV-------------- 385
```

Right side (second set of alignments):

```
Db    1146 STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205
           ||| ||  :| |  ||
Qy     386 ----------------------------------------------------------- 385
Db    1206 KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGLST 1265
Qy     386 ----------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
                                  :||||||:::| ::|::  ||||||::|||||||||
Db    1266 EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDTYTYTV 1325
                                  :|||||:| : :|:: |||| :|: |||| | |||||||
Qy     421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
           | : |:|||  |  :    :|:   |  |||||||||||||||:|||||| : :|
Db    1326 YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1384
Qy     481 TGSAVGQKVTLKWDAPN 497
           |||||||:|||||:|:|:
Db    1385 KAQPDGGDVLKWEAPS 1401

RESULT 8
AAY67396
ID   AAY67396 standard; protein; 1704 AA.
XX
AC   AAY67396;
XX
DT   25-APR-2000  (first entry)
XX
PN   Arg-gingipain-2 amino acid sequence.
XX
DE   Arg-gingipain-2 amino acid sequence.
XX
KW   Arginine specific proteinase; Arg-gingipain; gingipain-2; haemagglutinin;
KW   immunogenic component; vaccine; inflammatory response; tissue damage;
KW   periodontal disease.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key            Location/Qualifiers
FT   Region         229..719
FT                  /note= "Amino acids 229-719 are specifically claimed"
FT   Region         720..1185
FT                  /note= "Amino acids 720-1185 are specifically claimed"
XX
PN   US6017532-A.
XX
PD   25-JAN-2000.
XX
PF   08-NOV-1994;   94US-00336308.
XX
PR   10-SEP-1993;   93US-00119361.
PR   24-JUN-1994;   94US-00265441.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Potempa JS, Travis J;
XX
DR   WPI; 2000-136659/12.
DR   N-PSDB; AAZ60181.
XX
PT   New Porphyromonas gingivalis arginine-specific protease preparation
PT   useful for preparing vaccines against periodontal disease and for
PT   screening for Arg-gingipain inhibitors.
XX
PS   Claim 1; Col 29-42; 55pp; English.
XX
CC   This sequence represents a Porphyromonas gingivalis arginine-specific
CC   proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
CC   Gingipain-2 consists or a 50kD protease component non-covalently
CC   associated with a 44kD haemagglutinin component. The proteinase is
CC   stimulated by glycine containing peptides and glycine analogues. It is
CC   inhibited by cysteine protease group specific inhibitors. The protease
CC   preparation can be used in immunogenic compositions and vaccines against
CC   inflammatory response and tissue damage caused by P. gingivalis in
CC   periodontal disease. It can also be used to screen for agents that
```

```
OS  Porphyromonas gingivalis.
XX
FH  Key              Location/Qualifiers
FT  Peptide          1..227
FT                   /label= Prepro_peptide
FT  Region           599..619
FT                   /note= "Region of homology with cysteine proteases"
FT  Region           670..674
FT                   /label= Proteolytic_component
FT  Cleavage-site    719
FT  Protein          720..1091
FT                   /note= HGP 44kDa
FT                   /note= "Haemagglutinin protein component"
FT  Cleavage-site    1091
FT  Protein          1092..1429
FT                   /label= HGP 17kDa
FT                   /note= "Haemagglutinin protein component"
FT  Cleavage-site    1429
FT  Protein          1430..1704
FT                   /label= HGP 27kDa
FT                   /note= "Haemagglutinin protein component"
XX
PN  US6274718-B1.
XX
PD  14-AUG-2001.
XX
PF  25-JAN-2000; 2000US-00490931.
XX
PR  10-SEP-1993;   93US-00119361.
PR  24-JUN-1994;   94US-00265441.
PR  09-SEP-1994;   94WO-US010283.
PR  08-NOV-1994;   94US-00336308.
XX
PA  (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI  Travis J, Potempa JS, Barr PJ, Pavloff N;
XX
DR  WPI; 2001-588904/66.
DR  N-PSDB; AAS15242.
XX
PT  New recombinant DNA molecule which encodes high molecular weight (mature)
PT  Arg-gingipain protein, useful for immunization against inflammation and
PT  tissue damage, comprises enzymatically active protease component and
PT  hemagglutinin component.
XX
PS  Claim 1; Col 29-41; 56pp; English.
XX
CC  The invention relates to a recombinant DNA molecule encoding high
CC  molecular weight (mature) Arg-gingipain (AG) protein, which has an
CC  enzymatically active protease component (AG-2) and a haemagglutinin
CC  component, from P. gingivalis. The nucleic acid is useful for producing
CC  mature Arg-gingipain protein. Immunogenic compositions comprising Arg-
CC  gingipain are useful for immunising animals including humans against
CC  inflammatory response and tissue damage caused by an archaebacterium
CC  Porphyromonas gingivalis, which causes progressive periodontitis. Arg-
CC  gingipain is also useful for identifying agents that modulate Arg-
CC  gingipain proteinase activity, whether by acting on the proteinase itself
CC  or preventing the interaction of the proteinase with the protein in the
CC  gingival area, such as complement factors C3 or C5. The present sequence
CC  is Arg-gingipain-2
XX
SQ  Sequence 1704 AA;

Query Match          35.8%;   Score 944.5;   DB 4;   Length 1704;
Best Local Similarity 31.9%;   Pred. No. 4.4e-63;
Matches 235;   Conservative 75;   Mismatches 160;   Indels 267;   Gaps 20;

Qy    22 TAAAQG------GPKTAPSVTHQAVQK-GIR---TSKVKDLRDPIPAGMARIILEAH 68
Db   671 TATTQGGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730
Qy    69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPPEYKVPVNADASFSP 126
```

```
CC   modulate Arg-gingipain proteinase activity inhibitors
XX
SQ   Sequence 1704 AA;

Query Match          35.8%;   Score 944.5;   DB 3;   Length 1704;
Best Local Similarity 31.9%;   Pred. No. 4.4e-63;
Matches 235;   Conservative 75;   Mismatches 160;   Indels 267;   Gaps 20;

Qy    22 TAAAQG------GPKTAPSVTHQAVQK-GIR---TSKVKDLRDPIPAGMARIILEAH 68
Db   671 TATTQGGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy    69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPPEYKVPVNADASFSP 126
Db   731 DVWNDGSGYQILLDADHDQYGQVIPSDTHIW-PNCSVPANLFAFPEYTVPENADPSCSP 789

Qy   127 TNFVLDGTASADIPAGTYDVVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
Db   790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVPEAGKKYHFLMKKMG 849

Qy   186 PGDAASVV--
Db   850 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909

Qy   194 ------VTGEGSNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTYVLNESF 237
Db   910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESF 969

Qy   238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
Db   970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LITPDNYL 1025

Qy   298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS------ 350
Db  1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy   351 ---------DKPAPMNLV--------KSEGVKL------ 366
Db  1086 EAIRGRIQGTMRQKTVDLPAGTWTGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145

Qy   367 ----PAPYQERTID------------LSAYAGQQV-------- 385
Db  1146 STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy   386 --------------------- 386
Db  1206 KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST 1265

Qy   386 -------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
Db  1266 EANGAKPQSWIERTVDLPAGTKYVAFRHVNCSDLNYILLDDIQFTMGGSPTPTDYTYTV 1325

Qy   421 YRDNVVIAQNLAATFNQENVAPGQNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
Db  1326 YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1384

Qy   481 TGSAVGQKVTLKWDAPN 497
Db  1385 KAQPDGGDVVLKWEAPS 1401

RESULT 9
AAU08938
ID  AAU08938 standard; protein; 1704 AA.
XX
AC  AAU08938;
XX
DT  18-DEC-2001  (first entry)
XX
DE  P. gingivalis high molecular weight Arg-gingipain-2.
XX
KW  Periodontitis; antiinflammatory; Arg-gingipain-2; AG-2; immunogen.
XX
```

```
PI   Progulske-Fox A, Tummasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30656.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Clain 5; Page 138-143; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagE (AAR96033) was identified as the
CC   product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
CC   The haemagglutinin can be obtd. from transformed host cells and used as a
CC   vaccine to protect humans or animals against periodontal disease.
CC   Expression in Salmonella cells allows prod. of a live vaccine. The
CC   haemagglutinin can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1687 AA;

Query Match             35.7%;  Score 943.5;  DB 2;  Length 1687;
Best Local Similarity   31.8%;  Pred. No. 5.2e-63;
Matches  235;  Conservative  75;  Mismatches 160;  Indels 269;  Gaps 20;

Qy     22 TAAAQG-----------GPKTAPSVTHQAVQK--GIR--TSKVKDLRDPIPAGMARIILEAH 68
Db    652 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELRSGQAEIVLEAH 711

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
Db    712 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 770

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
Db    771 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 830

Qy    186 PGDAASVV------------------- 193
Db    831 SGDGTELTISEGGSGSDYTYTVYRDGTKIKEGLTATTFEBDGVATGNHEYCVEVKYTAGVS 890

Qy    194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS-- 235
Db    891 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTTLSE 950

Qy    236 SPDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
Db    951 SFE-NGIPASWKTIDADGDHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1006

Qy    296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVPLSTTGNEAANFTIKLLEETLGS--- 350
Db   1007 YLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEETITAKGVR 1066

Qy    351 ----DKPAPMNLV------------ KSEGVKL- 366
Db   1067 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1126

Qy    367 ------PAPYQERTID---------- 385
Db   1127 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL 1186

Qy    386 -------------------- LSAYAGQQV- 385
Db   1187 ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1246

Qy    386 -------------------- 386
Db   1247 STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLLDDIQFTMGGSPPTDYTY 1306

Qy    419 TVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ 478
Db   1307 TVYRDGTKIKEGLTETTFEBDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVK 1365
```

```
Db    731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789
Qy    127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
Db    790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy    186 PGDAASVV------------------- 193
Db    850 SGDGTELTISEGGSGSDYTYTVYRDGTKIKEGLTATTFEBDGVATGNHEYCVEVKYTAGVS 909

Qy    194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS-- DKRTYVLNESF 237
Db    910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTTLSESF 969

Qy    238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
Db    970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy    298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS------ 350
Db   1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy    351 -----------DKPAPMNLV------- KSEGVKL- 366
Db   1086 EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145

Qy    367 ----PAPYQERTID----- LSAYAGQQV- 385
Db   1146 STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy    386 -------------------- 385
Db   1206 KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST 1265

Qy    386 ------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
Db   1266 EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLLDDIQFTMGGSPPTDYTYTV 1325

Qy    421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
Db   1326 YRDGTKIKEGLTETTFEBDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVQNL 1384

Qy    481 TGSAVGQKVTLKWDAPN 497
Db   1385 KAQPDGGDVVLKWEAPS 1401
```

```
RESULT 10
AAR96033
ID   AAR96033 standard; protein; 1687 AA.
XX
AC   AAR96033;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagE.
XX
KW   Haemagglutinin; hagE; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain FDC381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995; 95WO-US016108.
XX
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
```

```
Qy    479 NLTGSAVGQKVTLKWDAPN 497
          ||   ||||:|:
Db   1366 NLKAQPDGGDVVLKWEAPS 1384

RESULT 11
AAW69495
ID   AAW69495 standard; protein; 1687 AA.
XX
AC   AAW69495;
XX
DT   22-DEC-1998  (first entry)
XX
DE   Haemagglutinin protein hagE.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagE.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;       95US-00570311.
XX
PR   08-SEP-1988;       88US-00241640.
PR   25-JAN-1991;       91US-00647119.
PR   09-DEC-1994;       94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58881.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 167-182; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagE haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1687 AA;

Query Match      35.7%;  Score 943.5;  DB 2;  Length 1687;
Best Local Similarity   31.8%;   Pred. No. 5.2e-63;
Matches 235;  Conservative 75;  Mismatches 160;  Indels 269;  Gaps 20;

Qy     22 TAAAQG------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
                        |||:  | :       : :||||||
Db    652 TATTQGGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQABIVLEAH 711

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
          |||:||:|::||||  | :   ::  |||   ||||||
Db    712 DVWNDGSGYQILLADHDQYGQVIPSDTHTIW-PNCSVPANLFAPFEYTVPENADPSCSP 770

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
          ||:|:::| |||||:||  : |:|  |||  :||::|| |
Db    771 TNMIMDGTASVNIPAGTYDFAIAAFQANAKIWIAGQGPTKEDDVVFEAGKKYHFLMKKMG 830

Qy    186 PGDAASVV------------ 193
                      ||
Db    831 SGDGTELTISEGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 890

Qy    194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTYVLNE 235
               || |||||||  ||    :  |   |  |   :|
```

```
Db    891 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTLSE 950
Db    236 SFDTQTLPNGWTMIDADGDHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
          | : :  ||
Db    951 SFE-NGIPASWKTIDADGDHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1006
Qy    296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
          ||:|| ::|||:||::   |||:::||:::|||
Db   1007 YLITPALDLIPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEETITAKGVR 1066
Qy    351 ------------DKPAPMNLV-----------KSEGVKL-- 366
                       : :                     |:|:
Db   1067 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1126
Qy    367 ------PAPYQERTID-------------LSAYAGQQV-- 385
                | :                      :|:
Db   1127 ESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNVL 1186
Qy    386 --------------------- 385
Db   1187 ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1246
Qy    386 --------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTY 418
                  :|||:|:|::|||  :|::|  |||  ||||
Db   1247 STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTY 1306
Qy    419 TVYRDNVVIAQNLAATTPNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ 478
          ||||| : ::   ::  | :|  :||||||||||||||||||  ::| :| : |:||| |
Db   1307 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVK 1365
Qy    479 NLTGSAVGQKVTLKWDAPN 497
          ||   ||||:|:
Db   1366 NLKAQPDGGDVVLKWEAPS 1384

RESULT 12
AAR96032
ID   AAR96032 standard; protein; 1358 AA.
XX
AC   AAR96032;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis hagD haemagglutinin.
XX
KW   Haemagglutinin; hagD; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain FDC381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;       95WO-US016108.
XX
PR   09-DEC-1994;       94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30655.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 125-129; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagD (AAR96032) was identified as the
CC   product of the second open reading frame of the hagD gene (AAT30655)
```

CC   derived from P. gingivalis 318 genomic DNA. A first open reading frame
CC   coded for hagD protease (see also AAR96031). The protease and
CC   haemagglutinin can be obtd. from transformed host cells and used in
CC   vaccines to protect humans or animals against periodontal disease.
CC   Expression in Salmonella cells allows prodn. of live vaccines. The
CC   haemagglutinin and protease can also be used to detect the presence of
CC   anti-P. gingivalis antibodies and to raise monoclonal antibodies for
CC   diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1358 AA;

Query Match   30.9%;   Score 816; DB 2; Length 1358;
Best Local Similarity 29.6%;   Pred. No. 2.4e-53;
Matches 219; Conservative 75; Mismatches 171; Indels 274; Gaps 21;

Qy   22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD-------PIPAGMARIILEA 67
                : ||| | | :::| |||||||||||
Db   328 TATTQGQKVTLKWDAPS-----AKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAA 382

Qy   68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
            :|||||||:|| | | |||| ||| ||| ||| |||| ||| | | |:||:||||
Db   383 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 441

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
                : | | |||: | |::| | |||:| ::| ||||||||||||
Db   442 QNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGDNQPARYDDFTFEAGKKYTFTM 501

Qy   182 QRQGPGDAASVV-------------------------------DKRTYVLNE 235
            :|||:||||||| :|||||||
Db   502 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 561

Qy   194 -----------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------------DKRTYVLNE 235
                       |||||| ||||||| || ||::| | |
Db   562 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTISE 621

Qy   236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
            :|| | ||:|| ||||| | ||:: | : | ::|| ||| ||| |||||
Db   622 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 677

Qy   296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
            ||:|: | |||:|:|||: |||||||| ||:|| ||| |||||| | :
Db   678 YLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEEFITAKGVR 737

Qy   351 ---------DKPAPMNLV----------------KSEGVKL------ 366
                     :|| |||:| |:|| ||
Db   738 SPEAIRGIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 797

Qy   367 -----PAPYQERTID-----------LSAYAGQQV----------- 385
                 |||:|| || | | | :|
Db   798 ESSTHGEAPAEWTTIDADGDGDWLCLSSGQLDWITAHGGTNVVASFSWNGMALNPDNYL 857

Qy   386 ----------- 385
Db   858 ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 917

Qy   386 -----------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTY 418
                             ||| |||||| || ||||| ||||||:||||||
Db   918 STEANGAKPQSVWIERTVDLPAGTKVVAFRHNCSDLDYILLDDIQFTMGGSPTPDYTY 977

Qy   419 TVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ 478
            ||||| | |:|: ||||:| ||| || | |:::||| | |: | ::| | ::
Db   978 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNTI-NPTQFNPVK 1036

Qy   479 NLTGSAVGQKVTLKWDAPN 497
            |||:|||||||||||:
Db   1037 NLKAQPDGGDVVLKWEAPS 1055

RESULT 13
AAW69494
ID   AAW69494 standard; protein; 1358 AA.
XX
AC   AAW69494;

XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagD.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagD.
OS   Porphyromonas gingivalis.
XX
PN   US582479l-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995; 95US-00570311.
XX
PR   08-SEP-1988; 88US-00241640.
PR   25-JAN-1991; 91US-00647119.
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UTFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58880.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 145-158; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagD haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1358 AA;

Query Match   30.9%;   Score 816; DB 2; Length 1358;
Best Local Similarity 29.6%;   Pred. No. 2.4e-53;
Matches 219; Conservative 75; Mismatches 171; Indels 274; Gaps 21;

Qy   22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD-------PIPAGMARIILEA 67
                : ||| | | :::| |||||||||||
Db   328 TATTQGQKVTLKWDAPS-----AKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAA 382

Qy   68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
            :|||||||:|| | | |||| ||| ||| ||| |||| ||| | | |:||:||||
Db   383 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 441

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
                : | | |||: | |::| | |||:| ::| ||||||||||||
Db   442 QNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGDNQPARYDDFTFEAGKKYTFTM 501

Qy   182 QRQGPGDAASVV------------------------------- 193
            :|||:|||||||
Db   502 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 561

Qy   194 VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------------DKRTYVLNE 235
            |||||| ||||||| || ||::| | | :|||||||
Db   562 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTISE 621

Qy   236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
            :|| | ||:|| ||||| | ||:: | : | ::|| ||| ||| |||||
Db   622 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 677

Qy   296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
            ||:|: | |||:|:|||: |||||||| ||:|| ||| |||||| | :
Db   678 YLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEEFITAKGVR 737

```
Qy   351  --------------DKPAPMNLV---------------KSEGVKL------  366
                        ||                        :  |
Db   738  SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF  797

Qy   367  ------PAPYQERTID-----------LSAYAGQQV-----  385
                :  ||                      :  :  |
Db   798  ESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL  857

Qy   386  -----------------------  385
Db   858  ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGINAGDFTVVFEETPNGINKGGARFGL  917

Qy   386  ----------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT  418
                                :||||:|:  :  :||:  : ||
Db   918  STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYILLDDIQFTMGGSPTPTDYTY  977

Qy   419  TVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ  478
          ||||:  :  :  :  ||  :  |||  :: ||||||||||||||: |::|||  :  :
Db   978  TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTV-NPTQFNPVK  1036

Qy   479  NLTGSAVGQKVTLKWDAPN  497
          ||  :  :|: ||||:||:
Db  1037  NLKAQPDGGDVVLKWEAPS  1055
```

RESULT 14
AAR96029
AAR96029;
ID   AAR96029 standard; protein; 1732 AA.
XX
AC   AAR96029;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis porphypain.
XX
KW   Porphypain; haemagglutinin; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain W12.
XX
FH   Key   Location/Qualifiers
FT   Region   688..708
FT   /note= "Pro-Asn repeat region type 1"
FT   Region   887..952
FT   /note= "Pro-Asn repeat region type 2"
FT   Region   946..967
FT   /note= "Pro-Asn repeat region type 1"
FT   Region   985..1006
FT   /note= "Pro-Asn repeat region type 3"
FT   Region   1041..1100
FT   /note= "Pro-Asn repeat region type 4"
FT   Region   1341..1405
FT   /note= "Pro-Asn repeat region type 2"
FT   Region   1430..1451
FT   /note= "Pro-Asn repeat region type 3"
FT   Region   1488..1547
FT   /note= "Pro-Asn repeat region type 4"
FT   Region   1607..1650
FT   /note= "Pro-Asn repeat region type 2"
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995; 95WO-US016108.
XX
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX

DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30653.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 76-81; 153pp; English.
XX
CC   P. gingivalis W12 cysteine protease, porphypain (AAR96029), was
CC   identified as the product of the prtP gene (AAT30653) isolated from P.
CC   gingivalis W12 genomic DNA. The porphypain shows homology to the
CC   haemagglutinins (see also AAR96026-28 and AAR96030-33) of P. gingivalis
CC   318. It can be obtd. from transformed host cells and used as a vaccine to
CC   protect humans or animals against periodontal disease. Expression in
CC   Salmonella cells allows prodn. of a live vaccine. The porphypain and
CC   haemagglutinins can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1732 AA;

Query Match   30.8%;   Score 812.5;   DB 2;   Length 1732;
Best Local Similarity   30.0%;   Pred. No. 6.4e-53;
Matches 222;   Conservative 70;   Mismatches 172;   Indels 277;   Gaps 22;

```
Qy    22  TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD----------PIPAGMARIILEA   67
                :  |                                            :
Db   693  TATTQGQKVTLKWEAPS---AKKAEGSREVKRIGDLFVTIEPANDVRANEAKVVLAA   747

Qy    68  HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP  126
          |: || ||:|| |||||| :  ||||  : |  || :  |:   |||| ||||||  : :
Db   748  DNVWGDNTGYFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVTT   806

Qy   127  TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV  181
             :  :|  :  ||| ||||  :||    :  :  :  ||:  : :   ||:  |  ||:
Db   807  QNIIVTGQGEVVIPGGVYDYCITNPEASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM  866

Qy   182  ORQGPGDAASVV----------------------  193
Db   867  RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT  926

Qy   194  ----------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP-----------ASDKRTVVLN  234
                    :|| ||||||||||  : ||  |: ||| ||:||
Db   927  AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS   986

Qy   235  ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD  294
          |||   | |   |:|||||||  |   :       ||||      :    |  :|: | |
Db   987  ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGLAGYNSNGCVYSESF-GLGGIGV-LTPD  1042

Qy   295  NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---  350
          ||| ||  : | || |  :||    :  :  :|| : : : ||:  |||   ||||
Db  1043  NYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV  1102

Qy   351  -------------DKPAPMNLV---------------KSEGVKL-----  366
                          :  ||                       :  |
Db  1103  RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET  1162

Qy   367  ------PAPYQERTID-----------LSAYAGQQV-----  385
                :  ||                      :  :  |
Db  1163  FESSTHGEAPAEWTTIDADGDGGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY  1222

Qy   386  ----------------------------  385
Db  1223  LISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGINAGDFTVVFEETPNGINKGGARFG  1282

Qy   386  ----------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT  417
                                :||||:|:  :  :||:  : ||
Db  1283  LSTEANGAKPQSVWIERTVDLPAGTKVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT  1342

Qy   418  YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV  477
          |||||:  :  :  :  ||  :  |||  :: ||||||||||||||: |::|||  :  :
Db  1343  YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTV-NSTQFNPV  1401
```

```
Qy    478  QNLTGSAV--GQKVTLKWDAP 496
           |||| ||:||
Db   1402  QNLTAEQAPNSMDAILKWNAP 1422
```

```
RESULT 15
AAW24787
ID   AAW24787 standard; protein; 1732 AA.
XX
AC   AAW24787;
XX
DT   17-OCT-2003  (revised)
DT   25-NOV-1997  (first entry)
XX
DE   PrtK antigenic protein complex.
XX
KW   Periodontal disease; cell surface protein; thiol protease; endopeptidase;
KW   PrtK; PrtK48; PrtK39; PrtK15; PrtK44; haemagglutinin; adhesin; therapy;
KW   diagnosis; vaccine; antigen.
XX
OS   Porphyromonas gingivalis; strain W50.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..228
FT                   /label= Pro-pro_peptide
FT   Cleavage-site   228..229
FT   Protein         229..737
FT                   /label= PrtK48
FT                   /note= "48 kDa Lys-specific thiol protease"
FT   Cleavage-site   737..738
FT   Protein         738..1156
FT                   /label= PrtK39
FT                   /note= "39 kDa adhesin"
FT   Cleavage-site   1156..1157
FT   Protein         1157..1291
FT                   /label= PrtK15
FT                   /note= "15 kDa adhesin"
FT   Cleavage-site   1291..1292
FT   Protein         1292..1732
FT                   /label= PrtK44
FT                   /note= "44 kDa adhesin"
XX
PN   WO9716542-A1.
XX
PD   09-MAY-1997.
XX
PF   30-OCT-1996;  96WO-AU000673.
XX
PR   30-OCT-1995;  95AU-00006275.
XX
PA   (UYME ) UNIV MELBOURNE.
PA   (VICT-) VICTORIAN DAIRY IND AUTHORITY.
XX
PI   Reynolds EC,  Bhogal PS,  Slakeski N;
XX
DR   WPI; 1997-272112/24.
DR   N-PSDB; AAT78851.
XX
PT   New antigenic protein complex from Porphyromonas gingivalis - comprising
PT   Arg- and lys- specific thiol endo-peptidase(s), used in the detection,
PT   prevention and treatment of periodontal disease.
XX
PS   Example 1; Fig 9b; 68pp; English.
XX
CC   A PrtR-PrtK cell surface protein of Porphyromonas ginivalis (PG)
CC   comprises a 300 kDa complex composed a 48 kDa lysine-specific thiol
CC   protease and 39, 15 and 44 kDa adhesins encoded by the prtK gene
CC   (AAT78851), and a 45 kDa arginine-specific thiol protease and 44, 15, 17
CC   and 27 kDa adhesins (see AAW24783) encoded by the prtR gene (AAT78850). A
CC   claimed antigenic complex comprises at least one multimeric protein
CC   complex of PrtR and PrtK each containing at least one adhesin domain, the
CC   complex having a mol.wt. of over 200 kDa, and preferably comprises all 9
CC   proteins of the PrtR-PrtK complex (see also AAW24780-85). It can be used
```

```
CC   in a claimed composition to elicit an immune response directed against
CC   PG, and in a claimed method of reducing the prospect of PG infection
CC   and/or severity of disease. Antibodies directed against the complex are
CC   claimed for use in treating PG infection. Unlike whole PG cells or other
CC   previously prepared antigens based on fimbriae or the capsule, the PrtR-
CC   PrtK complex or component parts are safe and effective antigens. (Updated
CC   on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1732 AA;

Query Match             30.8%;  Score 812.5; DB 2;  Length 1732;
Best Local Similarity   30.0%;  Pred. No. 6.4e-53;
Matches  222;  Conservative  70;  Mismatches 172;  Indels 277;  Gaps  22;

Qy     22  TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD-------------PIPAGMARILLEA 67
               || ||:             :    |     |:::|
Db    693  TATTQGQKVTLKWEAPS-----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy     68  HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
            :|:| |||||||:| ||      ||||| :|||
Db    748  DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVTT 806

Qy    127  TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
            ||  ||||  |  ||||| |||: |||        |   :    : :     ||:|  ||:
Db    807  QNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 866

Qy    182  QRQGPGDAASVV-------------------------------------------------- 193
            :|: ||
Db    867  RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy    194  ---------------VTGRGGNEFAPVQNLQWSVSGQTVTLTWQAP------------ASDKRTYVLN 234
                           |||:||||||||  ||| | ||| ||    ||           |:
Db    927  AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPNPGTTLS 986

Qy    235  ESPDTQTLPNGWTMIDADGDGHRWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
            ||  | ||||||||||||||||| ||||||||| ||| |||  ||:|  ||||  ||:||
Db    987  ESFB-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVISESF-GLGGIGV-LTPD 1042

Qy    295  NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS--- 350
            |||||| | | ||||  ||:|||  ||| || :|||||||||||| :| ||||||| |
Db   1043  NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYYASSTGNDASNFTNALLEETITAKGV 1102

Qy    351  ----------------------------DKPAPMNLV------------------KSEGVKL------ 366
                                        |||| ||:                   :|:
Db   1103  RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy    367  --------PAPYQERTID-----------------LSAYAGQQV----------------- 385
                    ||| ||:|                   |:|| ||:|
Db   1163  FESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy    386  ----------------------------------------------------------- 385
Db   1223  LISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFG 1282

Qy    386  --------------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
                                      :|| | |  |:| || |     || ::    ||
Db   1283  LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYT 1342

Qy    418  YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
            |||| |||||||||||| |||||||||||:||||||||||||||||||||||||||::|:
Db   1343  YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTV-NSTQFNPV 1401

Qy    478  QNLTGSAV--GQKVTLKWDAP 496
            |||| ||:||
Db   1402  QNLTAEQAPNSMDAILKWNAP 1422

Search completed: May 18, 2004, 11:32:40
Job time : 29.9726 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:29:09 ; Search time 9.14177 Seconds
                                      (without alignments)
                                      5229.534 Million cell updates/sec

Title: US-08-353-485-2
Perfect score: 2641
Sequence: 1 MRKLNSLFSLAVLLSLLCWG.........QNLTGSAVGQKVTLKWDAPN 497

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR 78:*
              1: pir1:*
              2: pir2:*
              3: pir3:*
              4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2641	100.0	2628	2	T28651	hemagglutinin A -
2	950.5	36.0	1526	2	S49763	gingipain R (EC 3.
3	944.5	35.8	1704	2	A55426	gingipain R (EC 3.
4	812.5	30.8	1732	2	T30836	lysine-specific cy
5	494.5	18.7	991	2	I40229	arginyl endopeptid
6	156	5.9	1231	2	A48490	endo-1,4-beta-xyla
7	153	5.8	5291	2	F90696	hypothetical prote
8	152	5.8	1234	2	S72640	endo-1,4-beta-xyla
9	152	5.7	1483	2	C97012	probably celluloso
10	147	5.6	1983	2	G86643	hypothetical prote
11	144.5	5.5	781	2	T36143	probable secreted
12	138.5	5.2	2468	2	A83412	hypothetical prote
13	137.5	5.2	635	2	S19011	endo-1,4-beta-xyla
14	137	5.2	2013	2	AI1489	probable peptidogl
15	135.5	5.1	1433	2	A36734	bacillopeptidase F
16	135	5.1	1655	2	E97835	hypothetical prote
17	132	5.0	1651	2	JC1340	outer membrane pro
18	131.5	5.0	725	2	A90255	hypothetical prote
19	129.5	4.9	1873	2	T30944	surface protein pr
20	129	4.9	940	2	AB1744	internalin protein
21	129	4.9	1349	2	AI1476	cell surface prote
22	128	4.8	1348	2	AH1115	cell surface prote
23	128	4.8	4936	2	AH2515	hypothetical prote
24	127.5	4.8	5188	2	B85547	probable RTX famil
25	127	4.8	2710	2	A37052	toxin A - Clostrid
26	126.5	4.8	908	2	AE2254	hypothetical prote
27	126.5	4.8	1148	2	S72635	exo-poly-alpha-gal
28	125.5	4.8	940	2	AD1374	internalin protein
29	125	4.7	1503	2	T18266	cycloinulo-oligosa
30	124.5	4.7	868	2	AF3204	autotransporter pr
31	124.5	4.7	875	2	H90371	proteinase [import
32	124	4.7	1616	2	T17884	S-layer protein -
33	124	4.7	2013	2	AD1129	probable peptidogl
34	123.5	4.7	208	2	AF1718	hypothetical prote
35	122	4.6	729	2	T35028	probable glycosyl
36	122	4.6	1217	2	F97177	alpha-glucosidase
37	121.5	4.6	1530	2	AH1396	peptidoglycan anch
38	121	4.6	1142	2	C97080	levanase/invertase
39	120.5	4.6	587	2	A55368	transferred entry
40	120.5	4.6	1448	2	AI2007	subtilase family p
41	120	4.5	364	2	JE0292	fungal stress prot
42	120	4.5	1939	2	D97316	probable S-layer p
43	119	4.5	699	2	A38368	chitinase (EC 3.2.
44	119	4.5	2205	2	T08615	aggregation factor
45	118.5	4.5	694	2	B84331	hypothetical prote

ALIGNMENTS

RESULT 1
T28651
hemagglutinin A - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C;Accession: T28651
R;Han, N.; Whitlock, J.; Progulske-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A;Title: The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381 contains four 1
A;Reference number: Z20494; MUID:97047672; PMID:8926061
A;Accession: T28651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2628 <HAN>
A;Cross-references: EMBL:U41807; NID:g1552410; PID:g1469916; PIDN:AAB17128.1
C;Genetics:
A;Gene: hagA

Query Match   100.0%; Score 2641; DB 2; Length 2628;
Best Local Similarity   100.0%; Pred. No. 2.1e-163;
Matches 497; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60

Qy   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120

Qy   121 DASFSPTNFVLDGTASADIPAGTYDVIINPNGITYIVGEGVSKGNDYVVEAGKTYHFT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 DASFSPTNFVLDGTASADIPAGTYDVIINPNGITYIVGEGVSKGNDYVVEAGKTYHFT 180

Qy   181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240

Qy   241 TLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300

Qy   301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360

Qy   361 SEGVKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 SEGVKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420

Qy   421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPVCKDVTVEGSNEFAHVQNL 480

```
Db  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480

Qy  481 TGSAVGQKVTLKWDAPN 497
            ||||||||||||||||
Db  481 TGSAVGQKVTLKWDAPN 497


RESULT 2
S49763
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis (fragment)
C;Species: Porphyromonas gingivalis
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997
R;Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.
submitted to the EMBL Data Library, November 1994
A;Description: Cloning, sequence analysis and expression in Escherichia coli of prpR1 of
A;Reference number: S49763
A;Accession: S49763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <ADD>
A;Cross-references: EMBL:X82680
C;Genetics:
A;Gene: prpR1
C;Keywords: cysteine proteinase; hydrolase

Query Match         36.0%;    Score 950.5;  DB 2;    Length 1526;
Best Local Similarity  31.8%;    Pred. No. 1.1e-53;
Matches 236;  Conservative 77;  Mismatches 154;  Indels 275;  Gaps 20;

Qy   22 TAAAQG------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
            ||  |      :  | ||| :||  |   |  ::||:||| ||  | ||| |||||
Db  671 TAITTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
            | |||| | ||| | || :|||| |   ::| ||||| ||| ||||||||| |||||
Db  731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy  127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
            || :|||||:|||||||||||| | || :|: | ||:| || ||| ||||:|| |:||
Db  790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy  186 PGDAASVV--------------------------------------------------- 193
            |  |:||
Db  850 SGDGTELTISEGGSGSDYIYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVS 909

Qy  194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS--------------DKRTYVLNE 235
                  ||||||||||| | |||  | ||| ||| |              |||:||||
Db  910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESF 969

Qy  236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
            || : | ||||||||||||||| :|  | | |:|||:|::|  :|:|| ||| ||:| |
Db  970 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1025

Qy  296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
            |::||| ||||||||| ||||| ||| |:|| ||| :|||:||:|||||||||:
Db 1026 YLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1085

Qy  351 -----DKPA------------------------------------------------- 354
         : |||
Db 1086 SPEAIRGRIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1145

Qy  355 ---------------------PMNLV 359
                                 |:|:
Db 1146 ESSTHGERATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALNPDNYL 1205

Qy  360 ----------------------------------------- 359
            :
Db 1206 ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGL 1265

Qy  360 --KSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDV--AVSGEGSSND 415
            :|||:||  | ::: |||| || |  ::|  |||:|:||:|| |::    | |||:|
Db 1266 STEADGAKPQSVWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSFTPTD 1322
```

RESULT 3
A55426
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
N;Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP;
C;Species: Porphyromonas gingivalis
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
C;Accession: A55426; D53113
R;Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr,
J. Biol. Chem. 270, 1007-1010, 1995
A;Title: Molecular cloning and structural characterization of the Arg-gingipain protein
A;Reference number: A55426; MUID:95138080; PMID:7836351
A;Accession: A55426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1704 <PAV>
A;Cross-references: GB:U15282; NID:g557067; PIDN:AAA69539.1; PID:g557068
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isola
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: D53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 228-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141694)
C;Keywords: cysteine proteinase; hydrolase

```
Query Match         35.8%;    Score 944.5;  DB 2;    Length 1704;
Best Local Similarity  31.9%;    Pred. No. 3.2e-53;
Matches 235;  Conservative 75;  Mismatches 160;  Indels 267;  Gaps 20;

Qy   22 TAAAQG------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
            ||  |      :  | ||| :||  |   |  ::||:||| ||  | ||| |||||
Db  671 TAITTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
            | |||| | ||| | || :|||| |   ::| ||||| ||| ||||||||| |||||
Db  731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy  127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
            || :|||||:|||||||||||| | || :|: | ||:| || ||| ||||:|| |:||
Db  790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy  186 PGDAASVV--------------------------------------------------- 193
            |  |:||
Db  850 SGDGTELTISEGGSGSDYIYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909

Qy  194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS--------------DKRTYVLNESF 237
                  ||||||||||| | |||  | ||| ||| |              |||:||||||
Db  910 PKVCKDVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730
```

```
Qy   22 TAAAQG-------GPKTAPSVTHQAVOK--GIR--TSKVKDLRDPIPAGMARIILEAH 68
```

```
Db  671 TAITTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
Db  731 DVWNDGSGYQILLDADHDYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy  127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
Db  790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy  186 PGDAASVV--------------------------------------------------- 193
Db  850 SGDGTELTISEGGSGSDYIYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909

Qy  194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS--------DKRTYVLNESF 237
Db  910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESF 969

Qy  238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
Db  970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy  298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS------ 350
Db 1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy  351 -------DKPAPMNLV---------- 366
Db 1086 EAIRGRIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145
```

```
Qy    22 TAAAQGGPKTI----APSVTHQAVQKGIRTSKVKDLRD----------------PIPAGMARIILEA  67
            ||||||||        :  :: :   :    : |:|:|                        ||||||
Db   693 TATTQGQKVTLKWEAPS----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA        747

Qy    68 HDVWEDGTGYQMLWDADHNQYGASIPEERSFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
            :|:||||||||     |||||   :||.:|  :: ||.::|:||     ||| ||| ||||
Db   748 DNWWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVTT  806

Qy   127 TNFVLDGTASADIPAGTYDVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
            :  |||:||||| |||  : ||| |:       ||:      |:: ||:  ||||||| 
Db   807 QNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 866

Qy   182 QRQGPGDAASVV------------------------------------ 193
            :: : :: :|                                        
Db   867 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy   194 --------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP------------ASDKRTYVLN 234
                          |||| |:| ::| ||| |: :|   | ||                |||:||||
Db   927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS 986

Qy   235 ESFDTQTLPNGWTMIDADGDGHRWLSTINVYNTATHTDGAMFSKSWTASGGAKIDLSPD 294
            ||||  | |||  |  :   | || : :  :||   |||  | :  :|| :   |: || 
Db   987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTFPD 1042

Qy   295 NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
            |||| :|:|  || || |||:||  |||| ||||:| : |||| || ||||||:: ||    
Db  1043 NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102

Qy   351 --------------------DKPAPMNLV----------------KSEGVKL-------- 366
                                : ||| |:                    : :|         
Db  1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy   367 -------PAPYQERTID------------------LSAYAGQQV-- 385
                   ||||| ||                      |||| ||    
Db  1163 FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy   386 -------------------------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
                                                 ||||||||||||| ||||||  |||| |||||||
Db  1223 LISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTVFEETPNGINKGGARFG 1282

Qy   386 -------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
                               ||||||||||||| ||||||  |||| |||||||
Db  1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYT 1342

Qy   418 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
            ||||||||||||||||||||||||||||||||||||||||||||||:: : :|::||:: |
Db  1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTV-NSTQFNFV 1401

Qy   478 QNLTGSAV--GQKVTLKWDAP 496
            ||| |:|   |||||||:||  
Db  1402 QNLTAEQAPNSMDAILKWNAP 1422
```

```
Qy   367 ---PAPYQERTID--------------LSAYAGQQV------------------------ 385
          :. ||||| ||                |||| ||                          
Db  1146 STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy   386 -----------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
                             |||||||||||| ||||||  |||| ||||||||||
Db  1206 KDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTVFEETPNGINKGGARFGLST 1265

Qy   386 ----------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
                            |||||||||||| ||||||  |||| ||||||||||
Db  1266 EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYYYTV 1325

Qy   421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
            ||| :| :|:|:  |||::  :||    :|| || |||| || |:|:|   ||::  ::| 
Db  1326 YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1384

Qy   481 TGSAVGQKVTLKWDAPN 497
            ||| |:|||||||  |  
Db  1385 KAQPDGDGDVVLKWEAPS 1401
```

RESULT 4
T30836
lysine-specific cysteine proteinase porphypain (EC 3.4.22.-) - Porphyromonas gingivalis
N;Alternate names: lysine-specific cysteine proteinase 1, 60K
C;Species: Porphyromonas gingivalis
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30836; T30837; T30526; A53113
R;Barkocy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulske-Fox, A.; Lantz,
J. Bacteriol. 178, 2734-2741, 1996
A;Title: Analysis of the prtP gene encoding porphypain, a cysteine proteinase of Porphyr
A;Reference number: Z20895; MUID:96213011; PMID:8631659
A;Accession: T30836
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1732 <BAR>
A;Cross-references: EMBL:U42210; NID:g1314325; PID:g1314326; PIDN:AAB06565.1
R;Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z20896
A;Accession: T30837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-795,'I',797-1389,'N',1391-1478,'Y',1480-1732 <SLA>
A;Cross-references: EMBL:U75366; NID:g2182811; PID:g2182812; PIDN:AAB60809.1
R;Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A;Title: IS195, an insertion sequence-like element associated with protease genes in Por
A;Reference number: Z20844; MUID:98298016; PMID:9632563
A;Accession: T30526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1350,'Y',1352-1363,'Y',1365-1447,'H',1449-1732 <LEW>
A;Cross-references: EMBL:AF017059; NID:g2738802; PID:g2738803; PIDN:AAC26523.1
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: A53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 229-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141690)
C;Genetics:
A;Gene: prtP; prtK
C;Keywords: cysteine proteinases; hydrolase

Query Match   30.8%;   Score 812.5; DB 2; Length 1732;
Best Local Similarity 30.8%; Pred. No. 1.3e-44;
Matches 222;   Conservative 70;   Mismatches 172;   Indels 277;   Gaps 22;

RESULT 5
I40229
arginyl endopeptidase - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40229
R;Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A;Title: Structural characterization of arginipain, a novel arginine-specific cysteine
A;Reference number: I40229; MUID:95168884; PMID:7864651
A;Accession: I40229
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-991 <RES>
A;Cross-references: GB:D26470; NID:g927644, g927645; PIDN:BAA05484.1; PID:g927645

Query Match   18.7%;   Score 494.5; DB 2; Length 991;
Best Local Similarity 25.5%; Pred. No. 2.9e-24;
Matches 124;   Conservative 41;   Mismatches 90;   Indels 231;   Gaps 7;

F;537,539/Binding site: substrate (Asp, Asn). #status predicted

Query Match   5.9%;   Score 156; DB 1;   Length 1231;
Best Local Similarity   20.4%;   Pred. No. 0.043;
Matches 92; Conservative   64; Mismatches 128; Indels 166; Gaps   23;

```
Qy    106 AGLYDPFEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSK 165
              ||:||: | ||:|| :||||  |: ||| |    :|||| :||:|||:|:||||
Db     22 ASLFSP-----PIRVFADDTNINLVSNG----DFESGTIDGWIKQGNPTLAVTTEQAIGQ  72

Qy    166 GNDYVVEAGKTYHFTVQRQGP-----------GDAASV-----VVTGE-GGNEFAPVQNLQ 209
              ||    || :|| :|  ||            :|| |      ::|   |||:| |::||
Db     73 YSMKVTGRTQTY------EGPAYSFLGKMQKGESYSVSLKVRLVSGQNSSNPLITVTMFR  126

Qy    210 WSVSGQTV-TLTWQAPASD-----------------------------KRTYVLNESFD  238
              :| ||   ||  ||||                              |||| |||:|
Db    127 EDDNGKHYDTIVWQKQVSEDSWTTVSGTYTLDYIGTLKTLYMYVESPDPTLEYIDDVVV  186

Qy    239 ---------------TQTLPNGWTMIDADGDGHNWLSTIN---VNTATHTGDGAMFSKS  280
                             : |||||  : |  |  |      ||   ||:||:|:|| :|
Db    187 ----------------SGWIGTGSSVVKAVVGVA-HSGDYSLLTTG  237

Qy    281 WTAS-GGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANF 339
              | |      |:|||||:   :   : ||   || | ||||:||||||:||:  ||: ||
Db    238 RTANWNGPSYDL----------TGKIVPGQYNVDFWVK-----------FVNGNDTEQIKA  278

Qy    340 TIKLLEETLGSDKPAPM-----------NLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFN 393
              |:|| ||||||| |            ||  |   || :||  |||||||||||||||| :|::
Db    279 TVKAT----SDKDNYIQVNDFANVNKGEWTEIKGSF--TLPVADYSGISIYVESQ--N  328

Qy    394 STGIFRLYLDDVAVSGEGSSNDYT-----------------YTYYRD  423
               ||||||| |:||:|||| || ||:                ||||:|
Db    329 PT--LEFYIDDFSVIGEISNNQITIQNDIPDLYSVFKDYFPIGVAVDPSRLNDADPHAQL  386

Qy    424 ------NVVIAQNLAATTFNQENVAPGQYNY 448
                     |  |||| || ||| |  | | |:
Db    387 TAKHPNMLVAEN---AMKPESLQPTEGNF  412
```

RESULT 7
F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain R
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: F90696
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90696
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5291 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33965.1; PID:g13360000; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0542

Query Match   5.8%;   Score 153; DB 2;   Length 5291;
Best Local Similarity   24.3%;   Pred. No. 0.52;
Matches 137; Conservative   70; Mismatches 202; Indels 154; Gaps   33;

```
Qy      5 NSLFSLAVLLSLLCWGQTAAAQG--GPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGMA-  61
              | ::|::  |:| || || :||  ||| :||:| |||:   :||::||||||: |||:
Db   3303 NNGYTLTATVSDLAGNLGSASKGVTVDTTAPVISFNTVAGDDVINNVEHIQAQIISGTAT 3362

Qy     62 ------RIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESF-WPANG--TIPAGLYDPF 112
                    |:  |  :|   ||  |    |::|: :| |  : | |  |   |:|| ||||
Db   3363 GAVAGDRLV-----VTTAGQQYVTSTDASGN-WSVGVPASVISGLADGVTISATITD-- 3414

Qy    113 EYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVE 172
               |||||||||||:||||||||||| ||| ||:||||||||||||||||||||| ||::||
```

Qy   22 TAAAQG--------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
              | ||| |||:||:| |||:| |: |||||||||||||:|||:|||
Db   671 TATTQQKVTLKWDAPSTKINATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEB--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
              |||||||||||||||||||||||||: : ||| ||||||||||||:||||||||||
Db   731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTIW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNPGI-YIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
              |||||||||||||||||||||||||||| :||||||||||||:|||||||||||||||:
Db   790 TNMIMDGTASVNIPAGTYDPFAIAAFQANAKIWIAQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy   186 PGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQTLPNG 245
              :: | :| |:|||| :||: ||| || |
Db   850 SGDGTELTIS--- 859

Qy   246 WTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVP 305
              |:| | :| ||: :| : :| || || :| || ||| |:| | : | |:
Db   860 -- 859

Qy   306 ENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVKSEGVK 365
              :| |||:||||:||||||||:||: ||| | |:| ||||| |
Db   860 -- 859

Qy   366 LPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTYYRDNV 425
                                                     |:|||::|||| |||||:
Db   860 -----------------------------------EGGGSDYTYTYYRDGT 875

Qy   426 VIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAV 485
               | || | | | || | | || || | ||| | | | |:||| ||:|
Db   876 KIKEGLTETTYRDAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVV 935

Qy   486 GQKVTL 491
              |:|:||
Db   936 GKTITV 941
```

RESULT 6
A48490
endo-1,4-beta-xylanase (EC 3.2.1.8) A precursor - Thermoanaerobacterium saccharolyticum
C;Species: Thermoanaerobacterium saccharolyticum
C;Date: 03-May-1994 #sequence_revision 11-Apr-1997 #text_change 16-Jul-1999
C;Accession: A48490
R;Lee, Y.E.; Lowe, S.E.; Henrissat, B.; Zeikus, J.G.
J. Bacteriol. 175, 5890-5898, 1993
A;Title: Characterization of the active site and thermostability regions of endoxylanase
A;Reference number: A48490; MUID:93388520; PMID:8376336
A;Accession: A48490
A;Molecule type: DNA
A;Residues: 1-1231 <LEE>
A;Cross-references: GB:M97882; NID:g533366; PIDN:AAA21812.1; PID:g533367
A;Experimental source: strain B6A-RI
A;Note: mutation experiments support roles for Asp-537, Glu-600, and Asp-602; His-572 ca
-495, Asn-539, and His-572 are supported by X-ray reference A55905 on another molecule w
C;Genetics:
A;Gene: xynA
C;Function:
A;Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans
A;Pathway: xylan degradation
C;Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology
a xylanase A cellulose-binding repeat homology
C;Keywords: duplication; glycosidase; heat-stable protein; hydrolase; polysaccharide deg
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-1231/Product: endo-1,4-beta-xylanase A #status predicted <MAT>
F;39-186/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA1>
F;196-340/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA2>
F;386-676/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
F;683-854/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC1>
F;858-1043/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC2>
F;1056-1109/Domain: S-layer repeat homology <SLR1>
F;1115-1168/Domain: S-layer repeat homology <SLR2>
F;1179-1231/Domain: S-layer repeat homology #status atypical <SLR3>
F;495,572,600,602/Active site: Glu, His, Glu, Asp #status predicted

```
Db  3415 ------SAGNSSTQTHNVQVNTAAVSLSVST----------------ISGDNLIN---AAE 3450
              :                  ||||| |:|             ::|||||     ||
Qy   173 AGKTYHFTVQRQGPGDAASVVVT---GEGGNEFAPVQ-NLQWSVSGGQTVTLTWQAPASD  227

Db  3451 AGSA--LTLSGTGTNFATGTVVTVLLNGKGYS-ATIQSNGSWSVN---VPAADVAALSD  3503
          || :      :|| | ||| |:           ||| ||| |||     |||| |:|:
Qy   228 KRTYVLNESFDTQTLPNGWTMIDADGDGHNWLSTINV-YNTA------THTGDGAMFSK   279

Db  3504 GTSYTVSASAQ-------DSAGNGNSSTQTHNVQVNTAAVSLSVSTISGDNLINA-      3551
          :|:|  :  ||        |:|| ||: |  |||  :|  : : :||   | :
Qy   280 SWTASGGAKIDLS--PDNYLVTPKVTVPENGK-LSYWVSSQVPWT----------NE    323

Db  3552 ---AEAGSALITLSGTGTNFATGTVVTVLLNGKGYSATIQSNGSWSVNVPAADVAALSDGT 3608
             : || |  :  ||| || |||  || | :    : |: |     ||| ||||
Qy   324 HYGVFLS---TTGNEAANFTIKLLEETLGSDKPAP---MNLVKSEGVKLPAPYQERTIDL   377

Db  3609 SYTVSASAQDSAGNSAT------ASRSVAVDLTAPVISINTVSTDD-RLNAAEQQQPLTL  3661
          ||||||:   :|| |:|       : |: | | |||:| |  |:|  |:|||||| |||
Qy   378 ----SAYAGQQY---------YLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTVYRDNV 425

Db  3662 NGSTSAEVGQTVTVTFGGKTYTATVAANGTWALNVPAVDLAALGQG----------      3707
          | ||:|  ||||| || |:|  |||   | : | | |||  | |
Qy   426 VIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTG---  482

Db  3708 -AQTTTASV-NDRAGNPGQATHALIVD---TVAPTV-TIATVAGDDIINNAEQLAGQTI   3760
           : |||:|  |  ||| ||||:|: :|    |||||  || || ||| || |||| |||
Qy   483 ----SAVGQKVTL-----KWDA                                       495

Db  3761 SGTTTAEVGQTVTVTFNGQTWSA                                      3783
```

RESULT 8
S72640
endo-1,4-beta-xylanase (EC 3.2.1.8) xynA precursor - Thermoanaerobacterium thermosulfuri
C;Species: Thermoanaerobacterium thermosulfurigenes
C;Date: 29-Jul-1997 #sequence revision 29-Jul-1997 #text_change 16-Jul-1999
C;Accession: S72640; S72622
R;Matuschek, M.; Sahm, K.; Bahl, H.
submitted to the EMBL Data Library, March 1996
A;Description: Characterization of genes from Thermoanaerobacterium thermosulfurigenes E
A;Reference number: S72640
A;Accession: S72640
A;Molecule type: DNA
A;Residues: 1-1234 <MATU>
A;Cross-references: EMBL:U50952; NID:g1255235; PIDN:g1255238
A;Experimental source: strain EM1
R;Matuschek, M.; Sahm, K.; Zibat, A.; Bahl, H.
Mol. Gen. Genet. 252, 493-496, 1996
A;Title: Characterization of genes from Thermoanaerobacterium thermosulfurigenes EM1 tha
A;Reference number: S72621; MUID:97033555; PMID:8879252
A;Accession: S72622
A;Molecule type: DNA
A;Residues: 815-1234 <MAW>
A;Cross-references: EMBL:U50952
C;Genetics:
A;Gene: xynA
C;Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;
a xylanase A cellulose-binding repeat homology
C;Keywords: glycosidase; hydrolase
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-1234/Product: endo-1,4-beta-xylanase XynA #status predicted <MAT>
F;38-185/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA1>
F;195-339/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA2>
F;385-675/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
F;682-853/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC1>
F;857-1042/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC2>
F;1055-1108/Domain: S-layer repeat homology <SLR>

```
Query Match             5.8%;  Score 152;  DB 2;  Length 1234;
Best Local Similarity   20.0%;  Pred. No. 0.078;
Matches    97;  Conservative   66;  Mismatches 138;  Indels 184;  Gaps 23;

Qy   118 VNADASFSPTNFVLDGTASADIPAGTYDYVIINPNGIIYIVGEGVSKGNDYVVEAGKTY   177
```

```
Db    28 IRAFADDTNTNLVSNG--------DFETGTIDGWIKQGNPTLEVTTEQAIGQYSMKVTGRTQTY  83
           :  |     |||| |         ||| | ||||  |  ||| || ||  || |:|  | | ||
Qy   178 HFTVQRQGP----------------GDAASV---VVTGE-GGNEFAPVQNLQWSVSGGQTV-TLT 220

Db    84 ------EGPAYSFLGKMQKGESYNVSLKVRLVSGQNSSNPFITVTMFREDDNGKHYDTIV      137
               || |:|  |     |       | |           |: | | ||:|:|||
Qy   221 WQAPASDKRTYVLNESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKS     280

Db   138 WQKQVSE-------DSWTTVSGTYTLDYTGT--LKTLYMY-------------            168
          |:|  |:        ::|| |||||| |||||  |||  |:
Qy   281 WTASGGAKIDLSPDNYLVTPKVTVP---------ENGKLSYWVSSQVPWTNEHYGV---     327

Db   169 -VESPDPTLEYYIDDVVTPQNPIQVGNVINGTFENGNTSGWVGTGSSVVKAVYGVAHS      227
           | |  |  :|| ||||  |  |  | |  || :|:|||| | :|   |:| |  :
Qy   328 -FLSTTGNEA-------------ANFTIKLL---EETLGSDKPAP---                355

Db   228 GGYSLLTTGRTANWNGPSYDLTGKIVPGGQYNVDFWVKFVNGNDTEQIKATVKATSNKDN    287
                    | :|     ||:| |   | |||| |  |    |: |
Qy   356 ---------MNLVKSEGVKLPAPYQERTIDLSAYAGQQVLAFRHFNSTGIFRLYLDDVAVS  408

Db   288 YIQVNDFVNVNKGEWTEIKGSF---TLPVTDYSGVSIVVESQ-NPT-LEFYIDDFSVI     340
           :: : |:|:|||| |||| |:|    |||:|:||| |||||:  ||: |||||||:||
Qy   409 GEGSSNDYT----------YTVYRD------NVVIAQNLAA                     433

Db   341 GEISNNQITIQNDIPDLYSVFKDYFPIGVAVDSSRLNDADPHAQLTAKHFNMLVAEN---    397
           ||:|||:|   ||||||| ||:|||| || ||     | | |||||||||||| |::
Qy   434 TTFNQENVAPGQYNYCVE-----VKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQK   488

Db   398 -AMKPESLQPTEGNFTFDNADKIVDY----------------------EIAHNMKMRGH---  433
           ::   :|:|  |||:|:|    |:|                        : :|
Qy   489 VTLKW                                                           493
```

```
Db   434 -TLLW                                                           437
```

RESULT 9
C97012
probably cellulosomal scaffolding protein precursor, secreted, cellulose-binding and col
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: C97012
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clc
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C97012
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1483 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78886.1; PID:g15023809; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0910

```
Query Match             5.7%;  Score 150;  DB 2;  Length 1483;
Best Local Similarity   20.2%;  Pred. No. 0.14;
Matches   127;  Conservative   81;  Mismatches 213;  Indels 208;  Gaps 29;

Qy    27 GGPKTAP--------SVTHQAVQKGIRTSKVKDLRDPIPAGMARIILEAHDVWEDGTGY-    77
              |       | |:  | |    :  ||  | :|        ||
Db   522 GTPVVTPSQINVEQGSATDQPVKIDLNGNTLKDVVDQ------SGKTLVQGTDYTVTDTGIT 577

Qy    78 -----------QMLWDADHRNQYGAS----------IPEESFWFANGTIPAGLYDPFEYKVPV 118
                       |:|  |: : ||  |            ||  ||  ||| || | | |  ||
Db   578 LSQSYLAGLALGQYTLLDFNGGGASQTITINVKNETVKLSVGTVSGNPGD--TVKVPV      635

Qy   119 N----------ADASFSPINF-VLDGTASADIPAGT--YDYVIINPNPGIIYIV---       159
          :          ||:||  :   |  :|:||  |:          : |  | |::
Db   636 TISQVSTPVGLICMDISYDASKFTVKDVLPNTDLVKDTDNYSFIVNTSTPGKISITFTDP   695

Qy   160 -------GEGVSKGNDYVVEAGKT--YHFTVQRQGPGDAASVVVTGEGGNEFAPVQNL     208
```

```
Db   825  SGSAKVYEATINKNGTYTLGAEVTSD--MGIKIVESKGSVKVEL-PDGSTKAYVLIFETS   881
                                     :  ::     :|   :  |:     |  | |:::|
Qy   304  VP----------------------ENGKLSYWVSSQVPWTNEHYGVFLSTTGNE------   335
          :
Db   882  LEGNLLINQKEYKNKATFTNKDISHDLSASV-TPAHQGEFVTKDGSQSSTDSNYVNWKLT   939

Qy   336  --AANFTIKLLEETLGSDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFN   393
             :: |    :                   |:                 |  :
Db   940  VNASQSTLKNVEVT--DNFSSNQYIVAKDILI-------YGYSIDASG----------N   979

Qy   394  STGIFRLYLDDVAVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVK   453
          : :                                          : :
Db   980  ITENKNILE---------QGKDYSVDIQTDNSTGAQTVKKFLSEINTA----YVVE-   1024

Qy   454  YTAGVSPKVCKDVT-----VEGSNE----------------FAHVQNLTGSAVGQK--VTLKWDAP   496
          |    ||       ::    :  :                 :  |          ||
Db   1025 YRALITSDKANDVVTNQAHITGDNEKIIEQDVEKDVPVTNHNGSANGSKGSVTLEKVGP   1083
```

RESULT 11
T36143
probable secreted proteinase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2003
C;Accession: T36143
R;Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A;Reference number: Z21598
A;Accession: T36143
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-781 <SEE>
A;Cross-references: EMBL:AL096852; PIDN:CAB51001.1; GSPDB:GN00070; SCOEDB:SCE19A.20c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE19A.20c
C;Superfamily: secreted virulence protease, InhA type

Query Match   5.5%;   Score 144.5;   DB 2;   Length 781;
Best Local Similarity   21.2%;   Pred. No. 0.13;
Matches 111;   Conservative 58;   Mismatches 164;   Indels 191;   Gaps 31;

```
Qy   24   AAQGGPKTAP-SVTHQAVQKGIRTSKVKDLRDPIPAG---MARIILEAHDVWE---DG   74
                              :    :             |||        |||
Db   205  ARYGSNKCDPDNCAWYAVQDGV-TAWIADQK--AAGRTDAQIKSQLTQYDQWDRYDFDG   260

Qy   75   TG-----YQMLWDADHNQYGASIPEESF---WFANGTIPAGLYDPFEYKVPV   118
                  |||||||||||||||||||          :||
Db   261  DGDFNEPDGYIDHFQIVHAGEDESAGGAQGEDAIWAHRWYAFGT-DAGATGPADNK---   316

Qy   119  NADASFSPTNFVLDGTASADIPAGTYDYVIINENPGI-IYIVGEGVSKGNDYVVEAGKTY   177
                           ||  :||     | |:     :: |             |
Db   317  ----------LGGTQIGDTGIWVGDYTIQPENGGLGVY------AHEYGHDLGLPD   356

Qy   178  HFTVQRQGPGDAASVVVTGEGGNEFAPVQNLQWSV--SGQTVTLTWQAPASDKRTYVLNE   235
                       |:                       :: |  ||
Db   357  HY---------DTAG-------GDNSTGF---------WTLMSSG-----SWLGTGR---NE   385

Qy   236  SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTDGAMFSKSWTASG---   285

Db   386  IGDLPGDMNAWDKLQL----GWLN----YDTA------KAGVNSWHKLGLAEYNTKHKQ   430

Qy   286  GAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTI---   341
                              :  ::|
Db   431  GLVVELPKEK--VTTEIVTPAEGETQWWSGSGNDLKNT-LSRSVDLTGKSAASLTLDGWY   487

Qy   342  KLLEETI------GSDKPAPMNLVKSEGVKLPAPYQE   372
             ::::
Db   488  DIEADYDFLYTEVSTDGGANWTAIDGTFDGNPIQRDGSDKPALSATVDAYG-KLJVIP-   543

Qy   373  RTIDLSAYAGQQVYLAFREFNSTGIFR--LYLDDVAVSGEGSSNDYTYTVYRDNV---   425
                 || || || |    | | |||       |||||||||||||||
Db   544  ----LDAYAGKKIDLRFRYQTDGGLAMKGFTADEIAVTADGE------TLFSDNAETADD   593
```

RESULT 10
G86643
hypothetical protein ybeF [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: G86643
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1983 <STO>
A;Cross-references: GB:AB005176; PID:g12723000; PIDN:AAK04249.1; GSPDB:GN00146
C;Genetics:
A;Gene: ybeF

Query Match   5.6%;   Score 147;   DB 2;   Length 1983;
Best Local Similarity   21.7%;   Pred. No. 0.32;
Matches 117;   Conservative 77;   Mismatches 181;   Indels 164;   Gaps 30;

```
Db   696  TLANYPISVDGILAYLDFIINSNATAGDSALTV------DPATHLVADEND----KDI   743
              :|: |    :| |:::::|               |::::|
Qy   209  QWSVSGQTVTLTWQAPASDKRTYVLNESFDT--QTLP------NGWTMIDA-DGDGH   256

Db   744  KDAASNGKITVTGSAPV--VQSSVVNTSSVTYDQNAPQDQAVSITFNGNTVKDVKDASG-   800
           : ::||||                :  ||
Qy   257  NWLSTINVYNTATHTDGAMFSKSWTASGGA----KIDLSPDN----YLVTPKVTVPE   306

Db   801  --NTLKAGSDYTATSDGITLSQSYLATLAAGTYTYTIDFSAGNAGTFTVVVKGKTVVGS   857
                                                   ||||::|   |:|
Qy   307  NGKLSYWVSS------QVPWT---NEHYGVFLSTTGNEAANFTIK----LLEET-   347

Db   858  ATTLAVGTVSGKAGDTVKPVTISKVTTPVGLICAEIDYDASKFTVKDVLPNTDLVKDTD   917

Qy   348  -----LGSDKPAPMNL---------VKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRH   391
                 |::::| |                    :                 | |:
Db   918  NYSFIVNTSTPGKISITFTDPTLANYPISADGI--LAYLDFIINSNATAGDSALTV--   971

Qy   392  FNSTGIFRLYLDDVAVSGEGSSNDYT------   417
               :::                 ::|
Db   972  -NPSGFIIADENDKDIQDAASNGKITVTGSTPVAENSVVNTSSVTYDQNAPQDQAVSITL   1030

Qy   418  ----------YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVK-T   455
                     ||||||||||||||||||||||||||||||||||
Db   1031 NGNTITDVKDASGNTLKAGSDYTVTSDGITLSQSYLAT-----LAAGTYTYTVDFSAGN   1084

Qy   456  AGVSPKVCKDVTVEGSNEFAHVQNLTGSA   484
          |||||||||||  ||||
Db   1085 AGTFTVVVKAKTVSSATTLAVGTVSGKA   1113
```

RESULT 12
A83412
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83412
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2468 <STO>
A;Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1874

Query Match   5.2%;   Score 138.5;   DB 2;   Length 2468;
Best Local Similarity   23.7%;   Pred. No. 1.6;
Matches 107;   Conservative 47;   Mismatches 166;   Indels 131;   Gaps   23;

Qy   103 TIPAGLYDPFEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEG 162
          | ||| ||| | : :| :: ||||::|:|:|:|||||||:|: |: :
Db   319 TDPAG-----NNSTPVTVEAP-------DTTA------PAPATD----------VQVAPDG 351

Qy   163 VS-KGNDYVVEAGKTYHFTVQRQFGPGDAASVVVTGEGNEFAPVQNLQWSVSGQTVTLTW 221
           | |||||||||| | :|||:||:: |||||||||||||||||||||| ||||||||||
Db   352 SSVTGN--AEPGAT--VGVDTDGDGQPDTTVVVGPGGSFEVPLN--PPLTNGETVTVIV 404

Qy   222 QAPASDKRTYVLNESFDTQTLPN-----------------GWTMIDADGDGH-------- 256
          |||||||||||||:||| | | |:: | || :
Db   405 TDPAGNSSTPVTAEAPDFPDAPQVNASNGSVLSGTAEAGVTIVIITDGNGNPIGQTSADAN 464

Qy   257 -NWLST------------INVY-------NTATH-TGDGAM------FSKSWTASG 285
          :|| | | :|| : | :|
Db   465 GNWSFTPGSQLPDGTVVNVVARDAAGNSSPATSITVDGVAPNAPVVEPSNGSELSGTAEP 524

Qy   286 GAKIDLSPDNVIVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLE 345
          || ||| ::| ||||||| :|||:|:|: |: || |:: |: :: |:|
Db   525 GSSVTLTDGNGNPIGQTTADANGNWSFTPSTPLP-----DGTVVNVVARDAAGNSSPPAS 579

Qy   346 ETLGSDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQOVVLAFRHFNSTGIFRLYLDDV 405
          :||||||| ||||||||||| |||||||||||||| |:|| |: |:|| |::
Db   580 -HNISVRG-ASSNAGTAKV---DLVIG----------GVTVGSFNF------ 621

Qy   406 AVSGEGSSNDYTYT---VYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPK 461
          | ::: || || | |:|:||||||||||||||||| ||||||||||||||||||||
Db   622 QVTADGSGN-WTFTPSTPLPNGTVV---NATATDPSGNASSPAS---VTVDAVAPATPV 673

Qy   462 V--CKDVTVEGSNEFAHVQNLT--GSAVGQ 487
          | :|| :|| | |:|:| |:|||
Db   674 VNPSNGTTLSGTAEPGATVTILTDGNGNPIGQ 704

RESULT 13
S19011
endo-1,4-beta-xylanase (EC 3.2.1.8) - Bacillus polymyxa
C;Species: Bacillus polymyxa
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 15-Oct-1999
A;Accession: S19011
R;Gosalbes, M.J.; Perez-Gonzalez, J.A.; Gonzalez, R.; Navarro, A.
J. Bacteriol. 173, 7705-7710, 1991
A;Title: Two beta-glycanase genes are clustered in Bacillus polymyxa: molecular cloning,

---

A;Reference number: S19011; MUID:92041687; PMID:1938968
A;Accession: S19011
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-635 <GOS>
A;Cross-references: EMBL:X57094; NID:g48815; PIDN:CAA40378.1; PID:g4816
A;Note: the authors translated the codon GAA for residue 78 as Gly, CCT for residue 272
C;Function:
A;Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans
A;Pathway: xylan degradation
C;Superfamily: Clostridium xylanase A repeat homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;408-502/Domain: Clostridium xylanase A repeat homology <CXA>

Query Match   5.2%;   Score 137.5;   DB 2;   Length 635;
Best Local Similarity   21.2%;   Pred. No. 0.27;
Matches 123;   Conservative 65;   Mismatches 176;   Indels 217;   Gaps   32;

Qy   22 TAAAQGGPKTAP---SVTHQAVQKGIRTSKVKDLRDPIPAGMARIILEAHDVWEDGTGY 77
          ||: | | : | :||: |:: || ::| :| ::| :|| ||| | :
Db   156 TADTPIGPWTDPLGKALVTHST---------------PGMAGVTWLFDPAVLVDDDGTGY 200

Qy   78 QMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKV----------------PVNADA 122
          | | : ||||: :|||| | | : :|| :: ||||
Db   201 --LYS-----GGGIPNES------DPASIANPKTARVIKLGADMTSVIGSATTIDAPY 245

Qy   123 SFSPT-----------------NFVLDGTASADIPAGTYDYVII-NP----------NPG 154
          |||| |:||||||||||||||||| :| |||
Db   246 LFEDSGIHKYNGKYYYSYCINFA--GTHQQYPAGEIGYMVSDNPMGPFTYKGHFLKNPY 303

Qy   155 IIYIVG-----------------EGVSKGNDYVVRAGKTY---HFT-VQRQGPGDAA 190
          ||| | ||| ||||||||||||| ||| |||:|||||
Db   304 TFFGVGGNNHHAVFNFKNEWYVVHAQTVSKAQ---IGAGKGYRSPHINKLVHKEDGSIS 360

Qy   191 SVV--VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQTLPNGWTM 248
          :|| :|||||||||||||||||||||||||: ||||||||||||||||||||| |
Db   361 EVQGNMTG------IAQLSMNPYTRVEAETIAWQAGVTTEPT------------------ 397

Qy   249 IDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAK---------------IDLSP 293
          |||| |||| |:||:|| ::|: || :|| | | ::
Db   398 -QASGGP---ISNLNVTN--IHNGDWIAVGKADFGSAGAKTFKANVATNVGGNIEVRLDS 451

Qy   294 DNYLVTPKVTVPENGKLSYW---VSSQVPWTNEHYGVFLSTTGNEAAN--------FTIKL 343
          ||:| ||| :||||:|| | |::|:| |:||||||||||||| || || ||
Db   452 ETGPIVGSLKVPSTGGMQTWREVETTINNATGVHNIYLVFTGSGSGNLLNLDAWQFTPNT 511

Qy   344 LEETLGSDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQOVVLAFRHF--NSTGIFRLY 401
          ::: ||||| |: |::||: |:: ||:|::|||:|||||||||||||| :::|||||
Db   512 GGNTITKVEAENMKIGGTYAGKISAPFD---GVALYANAD-YVSYSQYFANST--------- 560

Qy   402 LDDVAVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPK 461
          :||: :: :|| ||||||||||||||||||||||||||||||||||||||||||||||
Db   561 ----------GVTVGSFNF------------------------------TGKTPT 597

Qy   462 VCKDVTVEGSNEFAHVQNLTGSAVGQKVTL------KWDA 495
          | :||:: |:| | |||| : :||: ||
Db   598 V-----------QTLSNITHATGDQEIKLALTSDDGTWDA 626

RESULT 14
AI1489
probable peptidoglycan bound protein (LPXTG motif) lin0457 [imported] - Listeria innocu
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AI1489
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Brandt, P.; Chakraborty, T.; Charbit, A.; Chetouani, F.; Couve, E.; de Daruvar, A.;
., Dehoux, P.; Domann, E.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669

Qy   426 -----------------VIAQNLAATTFNQENVAPGQYNY 448
                           |||:||:||| :|||| |||
Db   594 AWTAVGFTRKGASFTKEYAQYIAENRQYVSYD-KTLKTGPYNF 636

A;Accession: AI1489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2013 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95689.1; PID:g16412898; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0457

Query Match   5.2%;   Score 137;   DB 2;   Length 2013;
Best Local Similarity   20.9%;   Pred. No. 1.5;
Matches   103;   Conservative   82;   Mismatches   163;   Indels   144;   Gaps   30;

```
Qy     74 GTGYQM-LWDADHNQYGASIPEES---FWFANGTIPAGLYDPFEYKVPVNADASFSPTNF 129
             :||  : :||||:    :  | ::: ||  :|:|
Db   1185 GQDYEIALFDKDGKEVQSAIKINNKGEFSFSD---VAIKNPADFKLKVTAP--TRTNF 1237

Qy    130 VLDGTASADIP---AGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGK-TYHFTVQRQG 185
             ||:|  ||:  :: |: :  |  ||  |::| :::| | :
Db   1238 VY----SAKNPLFNMSTKEYTLNSVVPGV------GGVAE--IYITETSKPTTKIILDKAV 1286

Qy    186 PGDAASVVVTGEGGNEFAPVQNLQWSV----------SGGQTV--            221
             | : : :::  |
Db   1287 TPNA----ITIESSDEATEVIN-EWTVVDSNGTVVYSGTGNTIRIPNDEGTYIAKNTATD 1341

Qy    222 QA--PASDKRTYVLNESFDTQTLPNGWTMIPADGDGH------NWLSTINVYNTATHTGD 273
             |: |:  |: :  :  |    :|  || :
Db   1342 EAGNTASDEKTFDIDY-IVPTLIVNQ------DASAEVNSTEANIDWIKPLNVAATDTHDGN 1396

Qy    274 ---GAMFSK-SW----------------IDLSPDNYLVIPKVTVPENG            308
               | | : |          :|  | :
Db   1397 ITPVVDYSKVKWDVLGTYPVTVTATDASGNKATQTVNLRIVDTTSPTILIT------NN 1449

Qy    309 KLSYWVSSQVPWTNEHYGVFLSTTGNE------------AANFTIKLLEETLG 349
             |: | :  :      ||:
Db   1450 PLTYSIENMRKLTEQELYKAAGLIGGDNYDLAPGQSVQPNKQPMVFTSNFS--TIFSDIA 1507

Qy    350 SDKP----APMNLVKSEGVK-LPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDD 404

Db   1508 SVKPGQYQVQVNLADSSGNQAIPQTITINVVDTMGPVIKADNVSY-HVNTTKTEAEFFQD 1566

Qy    405 VAVSGGRGSSNDYTYTVVRDNVVIAQNLAATTFNQENV--APGQYNYCVEVKYTAGVSPK 461
             || :| :  ||  :  :: :||  |: || : |
Db   1567 ARLDVTDNNDDTT------NLIITSNFA------EKVNLNKPGKYEVTISATDTKG--NQ 1612

Qy    462 VCKDVTVEGSNE 473
             |:|:| :| :
Db   1613 TTKEITVQVSKD 1624
```

RESULT 15
A36734
bacillopeptidase F (EC 3.4.21.-) precursor bpr [validated] - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000
R;Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
J. Bacteriol. 172, 1470-1477, 1990
A;Title: Bacillopeptidase F of Bacillus subtilis: purification of the protein and clonin:
A;Reference number: A35131; MUID:90170864; PMID:2106512
A;Accession: A35131
A;Molecule type: DNA
A;Residues: 1-365,'S',367-682,'EIMP',893,'Q',895-896 <SL2>
A;Cross-references: GB:M29035; NID:g143307; PIDN:AAA62679.1; PID:g143308
R;Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
J. Bacteriol. 172, 1470-1477, 1990
A;Title: Bacillopeptidase F of Bacillus subtilis: purification of the protein and clonin:
A;Reference number: A36734; MUID:90368623; PMID:2118514
A;Contents: erratum
A;Accession: A36734
A;Molecule type: DNA
A;Residues: 1-1433 <SLO>
A;Cross-references: GB:M29035; NID:g143035

R;Wu, X.C.; Nathoo, S.; Pang, A.S.H.; Carne, T.; Wong, S.L.
J;Biol. Chem. 265, 6845-6850, 1990
A;Title: Cloning, genetic organization, and characterization of a structural gene encod:
A;Reference number: A35750; MUID:90216713; PMID:2108961
A;Accession: A35750
A;Molecule type: DNA
A;Residues: 1-392,'V',394-828,'NIRTRLYSLKFCRSRHKSV' <WUA>
A;Cross-references: GB:J05400; NID:g142607; PIDN:AAA83362.1; PID:g142609
A;Note: this sequence has been corrected
A;Accession: B35750
A;Molecule type: DNA
A;Residues: 876-935,'CG','CG' <WU2>
A;Cross-references: GB:J05400; NID:g142607; PIDN:AAA83363.1; PID:g1119197
A;Note. this sequence has been corrected
R;Masuda, E.S.; Anaguchi, H.; Sato, T.; Takeuchi, M.; Kobayashi, Y.
Nucleic Acids Res. 18, 657, 1990
A;Title: Nucleotide sequence of the sporulation gene spoIIGA from Bacillus subtilis.
A;Reference number: S08223; MUID:90174995; PMID:2106671
A;Accession: S08223
A;Molecule type: DNA
A;Residues: 1410-1433 <MAS>
A;Cross-references: EMBL:X17344; NID:g40165; PIDN:CAA35224.1; PID:g809661
R;Kato, T.; Yamagata, Y.; Arai, T.; Ichishima, E.
Biosci. Biotechnol. Biochem. 56, 1166-1168, 1992
A;Title: Purification of a new extracellular 90-kDa serine proteinase with isoelectric]
A;Reference number: JN0335; MUID:93005071; PMID:1368833
A;Accession: JN0335
A;Molecule type: protein
A;Residues: 195-218,'A' <KAT>
A;Note: source of this material was Bacillus subtilis (natto)
R;Beall, B.; Lowe, M.; Lutkenhaus, J.
J. Bacteriol. 170, 4855-4864, 1988
A;Title: Cloning and characterization of Bacillus subtilis homologs of Escherichia coli
A;Reference number: I39846; MUID:89008108; PMID:3139638
A;Accession: I39849
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-211 <RES>
A;Cross-references: GB:M22630; NID:g142938; PIDN:AAA22458.1; PID:g951705
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte:
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch:
C.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fujita, M.; Fujita, M.; Fuma, S.; Galizzi, A.; Galle:
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F:
Koetter, F.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois:
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee:
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl T.M.; Portetell:
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon:
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero:
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama:
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,:
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69596
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1433 <KUN>
A;Cross-references: GB:Z99111; GB:Z99112; GB:AL009126; NID:g2633902; PIDN:CAB13404.1; P:
A;Experimental source: strain 168
C;Genetics:
A;Gene: bpr; bpf
A;Map position: 135 (degrees)
C;Superfamily: bacillopeptidase F; subtilisin homology
C;Keywords: extracellular protein; hydrolase; serine proteinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-194/Domain: propeptide #status predicted <PRO>
F;195-1433/Product: bacillopeptidase F #status experimental <MAT>
F;218-466/Domain: subtilisin homology <SBT>
F;227,274,452/Active site: Asp, His, Ser #status predicted

A;Residues: 1-365,'S',367-682,'EIMP',893,'Q',895-896 <SL2>
A;Cross-references: GB:M29035; NID:g143307; PIDN:AAA62679.1; PID:g143308
A;Note: the authors translated the codon GAA for residue 545 as Leu

Query Match   5.1%;   Score 135.5;   DB 1;   Length 1433;

Best Local Similarity 20.6%; Pred. No. 1.1;
Matches 112; Conservative 58; Mismatches 150; Indels 223; Gaps 31;

```
Qy   90  ASIPEESF---WFANGTIPAGLYDPFEYKV------PVNADASFSPTNFVLDGTAS------ 136
              :|:||: ||  :||   |||            |:|
Db  437  SSVPGQTYEDGW--DGTSMAG--PHVSAVAALLKQANASLSVDEMEDILTSTAEPLTDS     491

Qy  137  --ADIPAGTYDVIIINPNPGIIYIVGEGVSKGNDYVVEAGKT------YHFTVQRQGPGD   188
             |: ::|: : :|:|              |
Db  492  TFPDSPNNGYGHGLVNAFDA-VSAVTDGLGKAEGQVSVEGDDQEPPVYQHEKVTEAYEGG  550

Qy  189  AASVVVTGEGGNEFAPVQ-----NLQWS------VSGQTVILTWQAPASD----KRTY-  231
         : : :|  :|             ||:   |: |:||          :|: :|
Db  551  SLPLTLTAEDNSVTSVKLSYKLDQGEWTEITAKRISGDHLKGTYQAEIPDIKGTKLSYK  610

Qy  232  -----------VLNESFDT------------QTLPNGWTMIDADGDGHNWLSTINVYNT  267
                      :|:|           |||  :|:|  || :||
Db  611  WMIHDFGGHVVSSDVYDVTVKPSITAGYKQDFETAPGGWV---ASGTNNNW-------  658

Qy  268  ATHTGDGAMFSKSW-------TASGGAKI---DLSPDNY------LVTPKVTVPENGK  309
             ||:|           |||:  |:  ||  ::       |:::|
Db  659  -------------EWGVPSTGPNTAASGEKVYGTNLT-GNYANSANMNLVMPPIKAPDSGS  705

Qy  310  LSYWVSSQVPWTNEH----YG-VFLSTTG----NEAANFTIKLLEETLGSDKPAPMNLVK  360
               :  |||           |: ||
Db  706  LFLQFKS--WHNLEDDFDYGYVFVLPEGEKNWEQAGVYNGK----------------  744

Qy  361  SEGVKLPAPYQERTIDLSAYAGGQVYLAFRHFNSTGIFR--LYLDDVAVSGEGSSNDYTY  418
                    : :|    |||||||         |:  | | :|||||:|  |||:
Db  745  ------TSSWTDEIDLSAYKGGNIQVMFNLQSDESIAKEGWYIDDVVLSDKSAGK---  794

Qy  419  TVVRDNVVI-----------------------AQNLAATTFNQENVA----------  442
         ||||| |::           :|                :|
Db  795  TVKKNKLGVEKPSGKQKKPVNPKKAKPSANTAVKHQNKAIQPQVLPLKAQVSVVETGKS  854

Qy  443  -----PGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVG------QKVTL  491
               | |:|            ||:|  :| | ||      :|                |:|
Db  855  TYSDQSTGQYT----LKHKAG------DYT----------LMAEAYGYQSKTQKVSL   891

Qy  492  KWD 494
         |||
Db  892  KTD 894
```

Search completed: May 18, 2004, 11:35:56
Job time : 13.1418 secs

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.


OM protein - protein search, using sw model


Run on:     May 18, 2004, 11:22:23 ; Search time 6.24316 Seconds
                                   (without alignments)
                        4145.156 Million cell updates/sec

Title:           US-08-353-485-2
Perfect score:   2641
Sequence:        1 MRKLNSLFSLAVLLSLLCWG..........QNLTGSAVGQKVTLKWDAPN 497

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:        141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*


Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Accession	Description
1	2641	100.0	2628	1	HGA2_PORGI	Q51845	porphyromon
2	2583	97.8	2164	1	HGA1_PORGI	Q59915	porphyromon
3	494.5	18.7	991	1	CPG1_PORGI	P28784	porphyromon
4	472	17.9	989	1	PRTH_PORGI	P46071	porphyromon
5	156	5.9	1157	1	XYNA_THESA	P36917	thermoanaer
6	137.5	5.2	635	1	XYND_PAEPO	P45796	paenibacill
7	135.5	5.1	1433	1	SUBF_BACSU	P16397	bacillus su
8	135	5.1	1655	1	OMPB_RICCN	Q9kka3	r outer mem
9	131.5	5.0	725	1	YA33_SULSO	Q97z97	sulfolobus
10	131	5.0	1656	1	OMPB_RICJA	O06653	r outer mem
11	129	4.8	794	1	OXAA_CHLCV	P59809	chlamydophi
12	127	4.8	2710	1	TOXA_CLODI	P16154	clostridium
13	124	4.7	1616	1	SLAP_BACCI	P35824	bacillus ci
14	121.5	4.6	972	1	CTA1_BACCI	P94286	bacillus ci
15	120.5	4.6	587	1	PICP_PSSSR	P42790	pseudomonas
16	119	4.5	699	1	CHI1_BACCI	P20533	bacillus ci
17	118.5	4.5	1300	1	120K_RICRI	P14914	rickettsia
18	118.5	4.5	1654	1	OMPB_RICRI	Q53047	r outer mem
19	115.5	4.4	336	1	G3P1_SCHPO	P78958	schizosacch
20	115	4.4	335	1	G3PC_GRAVE	P54270	gracilaria
21	114.5	4.3	488	1	PHB_ALCFA	P12625	alcaligenes
22	114.5	4.3	681	1	TKT2_YEAST	P33315	saccharomyc
23	114.5	4.3	857	1	GELA_DICDI	P13466	dictyosteli
24	114.5	4.3	2660	1	YEEJ_ECO57	Q8x8v7	escherichia
25	113.5	4.3	1004	1	SLPO_BACBR	P09333	bacillus br
26	113.5	4.3	1151	1	ITA1_HUMAN	P56199	homo sapien
27	111	4.2	2201	1	TENA_HUMAN	P24821	homo sapien
28	111	4.2	2358	1	YEEJ_ECOLI	P76347	escherichia
29	110.5	4.2	408	1	G3PT_HUMAN	O14556	homo sapien
30	110.5	4.2	2265	1	FINC_BOVIN	P07589	bos taurus
31	110	4.2	642	1	FLID_CAMJE	Q9phw6	campylobact
32	110	4.2	827	1	CSG_HALVO	P25062	halobacteri
33	110	4.2	1409	1	HAP1_HAEIN	P44596	haemophilus
34	109.5	4.1	932	1	CDG3_HUMAN	Q9y5h0	homo sapien
35	109.5	4.1	2021	1	OMPA_RICCN	Q52657	rickettsia
36	109	4.1	591	1	FLGE_CAUCR	P35806	caulobacter
37	109	4.1	917	1	SLAP_THETH	P35830	thermus the
38	109	4.1	1120	1	STFR_ECOLI	P76072	escherichia
39	108.5	4.1	337	1	G3P_PODAN	P32637	podospora a
40	108	4.1	611	1	WDRI_CAEEL	Q11176	caenorhabdi
41	108	4.1	828	1	MRKC_KLEPN	P21647	klebsiella
42	108	4.1	2481	1	FINC_XENLA	Q91740	xenopus lae
43	107.5	4.1	1475	1	APU_THETY	P16950	t amylopull
44	107.5	4.1	2477	1	FINC_RAT	P04937	rattus norv
45	107	4.1	666	1	NEP1_THEVU	Q60053	thermoactin

ALIGNMENTS

```
RESULT 1
HGA2_PORGI
ID   HGA2_PORGI     STANDARD;      PRT;   2628 AA.
AC   Q51845;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97047672; PubMed=8926061;
RA   Han N., Whitlock J., Progulske-Fox A.;
RT   "The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381
RT   contains four large, contiguous, direct repeats.";
RL   Infect. Immun. 64:4000-4007(1996).
CC   -!- FUNCTION: Agglutinates erythrocytes.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U41807; AAB17128.1; -.
DR   PIR; T28651; T28651.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT   SIGNAL         1     24
FT   CHAIN         25   2628       HEMAGGLUTININ A.
FT   DOMAIN        25    539       PEPTIDASE C25-LIKE 1.
FT   DOMAIN       540    995       PEPTIDASE C25-LIKE 2.
FT   DOMAIN       996   1451       PEPTIDASE C25-LIKE 3.
FT   DOMAIN      1452   1907       PEPTIDASE C25-LIKE 4.
FT   DOMAIN      2074   2628       PEPTIDASE C25-LIKE 5.
SQ   SEQUENCE   2628 AA;  283324 MW;  61C4DE32540C99DA CRC64;

Query Match             100.0%;  Score 2641;  DB 1;  Length 2628;
Best Local Similarity   100.0%;  Pred. No. 1.1e-162;
Matches  497;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60

Qy   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
```

```
FT   DOMAIN      992   1443     PEPTIDASE C25-LIKE 3.
SQ   SEQUENCE   2164 AA;  233387 MW;  6DFAB22832586C63 CRC64;

Query Match          97.8%;  Score 2583;  DB 1;  Length 2164;
Best Local Similarity  98.0%;  Pred. No. 4.9e-159;
Matches  487;  Conservative  0;  Mismatches  10;  Indels  0;  Gaps  0;

Qy    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM  60
         |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKAKDLRDPIPAGM  60

Qy   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120

Qy  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180

Qy  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVLTWQAPASDKRTYVLNESFDTQ  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVLTWQAPASDKRTYVLNESFDTQ  240

Qy  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300

Qy  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDKPAPMNLVK  360
         | |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db  301 KFTVPENGKLSYWVSSQEPWTNEHYGVFLSTGNEAANFTIKLLEETLGSGKPAPMNLVK  360

Qy  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
         ||||| |||||||||||||||||||||||||    ||||||||||||||||||||||||
Db  361 SEGVKAPAPYQERTIDLSAYAGQQVYLAFRHFGCTGIFRLYLDDVAVSGEGSSNDYTYTV 420

Qy  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
         |||||||||| ||||||| ||||||||||||||||||||||||||||||||||||| |||
Db  421 YRDNVVIAQNLTATTFNQ2NVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNL 480

Qy  481 TGSAVGQKVTLKWDAPN 497
         |||||||||||||||||
Db  481 TGSAVGQKVTLKWDAPN 497
```

RESULT 2
HGA1_PORGI
```
ID   HGA1_PORGI     STANDARD;      PRT;   2164 AA.
AC   P59915;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA OR PG1837.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=22829867; PubMed=12949112;
RA   Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA   Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
RA   Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
RA   Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA   Dewhirst F.E., Fraser C.M.;
RT   "Complete genome sequence of the oral pathogenic bacterium
RT   Porphyromonas gingivalis strain W83.";
RL   J. Bacteriol. 185:5591-5601(2003).
CC   -!- FUNCTION: Agglutinates erythrocytes (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   ---------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------
DR   EMBL; AE017178; AAQ66831.1; ALT_INIT.
DR   TIGR; PG1837; -.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
KW   Complete proteome.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   2164       HEMAGGLUTININ A.
FT   DOMAIN       26    539       PEPTIDASE C25-LIKE 1.
FT   DOMAIN      540    991       PEPTIDASE C25-LIKE 2.
```

RESULT 3
CPG1_PORGI
```
ID   CPG1_PORGI     STANDARD;      PRT;    991 AA.
AC   P28784; Q45168;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
DE   (RGP-1).
GN   RGPA OR RGP1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 228-290 AND 517-541.
RC   STRAIN=381;
RX   MEDLINE=95168884; PubMed=7864651;
RA   Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
RA   Ikehara Y.;
RT   "Structural characterization of argingipain, a novel
RT   arginine-specific cysteine proteinase as a major periodontal
RT   pathogenic factor from Porphyromonas gingivalis.";
RL   Arch. Biochem. Biophys. 316:917-925(1995).
RN   [2]
RP   SEQUENCE OF 228-270.
RC   STRAIN=HG66;
RX   MEDLINE=92406812; PubMed=1527017;
RA   Chen Z., Potempa J., Polanowski A., Wikstrom M., Travis J.;
```

```
RT   "Purification and characterization of a 50-kDa cysteine proteinase
RT   (gingipain) from Porphyromonas gingivalis.";
RL   J. Biol. Chem. 267:18896-18901(1992).
CC   -!- FUNCTION: Thiol protease which is believed to participate in
CC       intracellular degradation and turnover of proteins. Its
CC       proteolytic activity is a major factor in both periodontal tissue
CC       destruction and in bacterial host defense mechanisms. Activates
CC       complement C3 and C5.
CC   -!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
CC       immunoglobulins, with a preference for Arg in P1, and hydrophobic
CC       residues in P2 and P3.
CC   -!- ENZYME REGULATION: Requires cysteine for activation and Ca(2+)
CC       and/or Mg(2+) for stabilization. It is stimulated by glycine-
CC       containing dipeptides. It is resistant to inhibition by proteinase
CC       inhibitors in human plasma.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
```

```
DR   EMBL; D26470; BAA05484.1; -.
DR   PIR; I40229; I40229.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
KW   Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
FT   SIGNAL        1     24
FT   PROPEP       25    227
FT   CHAIN       228    991       GINGIPAIN R1.
FT   ACT_SITE    438    438       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    471    471       NUCLEOPHILE (BY SIMILARITY).
FT   CONFLICT    264    265       RT -> TK (IN REF. 2).
SQ   SEQUENCE   991 AA;  108782 MW;  03E3F43CEBE2544 CRC64;
```

```
Query Match             18.7%;  Score 494.5;  DB 1;  Length 991;
Best Local Similarity   25.5%;  Pred. No. 2e-24;
Matches  124;  Conservative  41;  Mismatches  90;  Indels  231;  Gaps   7;

Qy     22 TAAAQG-------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH  68
             ||||             :|  ||  .  |::|    .: :  |        ||||||
Db    671 TATTGQQKVTLKWDAPSTKTNATTNTARSVDGIREJLVLLSVSDAPELLRSGQAEIVLEAH 730

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
           || .|||| |:|||||| |||: | |   ::|:  |:|::|   :||| ||| || ||||
Db    731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPPEYTVPENADPSCSP 789

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
             ||: |||::|::| |||| |:||| |||   |:  |::   ||| || |:  .  |:::
Db    790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQQPTKEDDYVFEAGKKTHFLMKKMG 849

Qy    186 PGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTVLNESFDTQTLPNG 245
              :      ::
Db    850 SGDGTELTIS------------------------------------------------ 859

Qy    246 WTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVP 305
Db    859 ------------------------------------------------------------ 859

Qy    306 ENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVKSEGVK 365
Db    859 ------------------------------------------------------------ 859

Qy    366 LPAPYQERTIDLSAYAGQQVVLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTVRDNV 425
            || :||||||
Db    425 ------------------------------------------MALNPDNYLISKDVTGATKVKYYYA 172
```

```
Db    860 -----------------------------------------------EGGGSDYTTVYRDGT 875

Qy    426 VIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAV 485
            || ::|  |  |:|:|:||| :|:|||  ||||:|| | |||  :.|::.|:  |  ||
Db    876 KIKEGLTETTYRDAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTV 935

Qy    486 GQKVTL 491
           |.: :|:
Db    936 GKTITV 941
```

```
RESULT 4
PRTH_PORGI
ID   PRTH_PORGI     STANDARD;      PRT;    989 AA.
AC   P46071;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protease prtH (EC 3.4.22.-).
GN   PRTH.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=95012612; PubMed=7927685;
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RT   "Cloning and characterization of a new protease gene (prtH) from
RT   Porphyromonas gingivalis.";
RL   Infect. Immun. 62:4279-4286(1994).
RN   [2]
RP   ERRATUM.
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RL   Infect. Immun. 62:5707-5707(1994).
CC   -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
CC       P.GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
CC       IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
CC       AND IS A VIRULENCE FACTOR.
CC   -!- SUBCELLULAR LOCATION: In membrane vesicles
CC   -!- SIMILARITY: Belongs to peptidase family C25.
```

```
DR   EMBL; L27483; AAA51298.1; -.
DR   HSSP; P23882; 1FMT.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR002376; formyl_transf.
KW   Pfam; PF00551; formyl_transf; 1.
KW   Hydrolase; Thiol protease; Repeat; Virulence.
FT   REPEAT       270    323
FT   REPEAT       528    581
SQ   SEQUENCE   989 AA;  110238 MW;  FA85FE8A3AC8944C CRC64;
```

```
Query Match             17.9%;  Score 472;  DB 1;  Length 989;
Best Local Similarity   38.6%;  Pred. No. 5.7e-23;
Matches 119;  Conservative  45;  Mismatches 124;  Indels  20;  Gaps  10;

Qy    199 GNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQT---LPNGWTMIDADGDG 255
Db     58 GTEICCPPSLPKAPICSTSTLMRLRSKTNAKRADFTETFBSSTHGRAPAEWTTIDADGDG 117

Qy    256 HNW--LSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYW 313
           |:||
Db    118 QGWLCLSSGQLDWLTAHGGTNVVSSFSWNG-----MALNPDNYLISKDVTGATKVKYYYA 172
```

```
Qy    314  VSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDK-PAPMNL-VKSEGVKLPAPYQ 371
           |:  :    ||| |:|:||   ||| |       :  :|||   :  ||:|
Db    173  VNDGFP--GDHYAVMISKTGTNAGDFTVFETPNGINKGGARFGLSTEADGAKPQSVWI 230

Qy    372  ERTIDLSAYAGGQVYLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTVYRDNVVIAQ 429
           |||||:| || | |:|:|  |:| ||: :|||   ||||||||||||||||||||||:
Db    231  ERTVDLP--AGTK-YVAFRHYNCSDLNVILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKE 287

Qy    430  NLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQKV 489
           |||| || |:    |||||   |||||||||  |||  : |||   |: :|:||  |:|
Db    288  GLTETTFEDGVATGNHEYCVBVKYTAGVSPKKCVNVTV-NSTQFNPVKNLKAQPDGGDV 346

Qy    490  TLKWDAPN 497
           |||:|:|:
Db    347  VLKWEAPS 354
```

RESULT 5

```
XYNA_THESA
ID   XYNA_THESA     STANDARD;       PRT;   1157 AA.
AC   P36917;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) (Xylanase A).
DE   XYNA.
GN   XYNA.
OS   Thermoanaerobacter saccharolyticum.
OC   Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC   Thermoanaerobacteriaceae; Thermoanaerobacterium.
OX   NCBI_TaxID=28896;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 34-41.
RC   STRAIN=B6A-RI;
RX   MEDLINE=94028940; PubMed=8215382;
RA   Lee Y.-E., Lowe S.E., Zeikus J.G.;
RT   "Gene cloning, sequencing, and biochemical characterization of
RT   endoxylanase from Thermoanaerobacter saccharolyticum B6A-RI.";
RL   Appl. Environ. Microbiol. 59:3134-3137(1993).
RN   [2]
RP   ACTIVE SITE, AND MUTAGENESIS.
RX   MEDLINE=93388520; PubMed=8376336;
RA   Lee Y.-E., Lowe S.E., Henrissat B., Zeikus J.G.;
RT   "Characterization of the active site and thermostability regions of
RT   endoxylanase from Thermoanaerobacterium saccharolyticum B6A-RI.";
RL   J. Bacteriol. 175:5890-5898(1993).
CC   -!- FUNCTION: Endo-acting enzyme that randomly cleaves the internal
CC       xylosidic linkages of the xylan backbone, yielding
CC       xylooligosaccharides of various lengths which are further
CC       hydrolyzed to xylose molecules by beta-xylosidase (EC 3.2.1.37).
CC       The temperature and pH optima for this enzyme are 70 degrees
CC       Celsius and 5.5, respectively. Requires at least three xylose
CC       residues for catalytic activity. Does not have activity against
CC       xylobiose.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
CC       linkages in xylans.
CC   -!- PATHWAY: Xylan degradation.
CC   -!- INDUCTION: By xylan and xylose.
CC   -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC       hydrolases).
CC   -!- SIMILARITY: Contains 2 S-layer homology (SLH) domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M97882; AAA21812.1; ALT_SEQ.
DR   HSSP; P10478; 1XYZ.
DR   InterPro; IPR008960; CBD9-like.
DR   InterPro; IPR003305; CBM_CenC.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR001000; Glyco_hydro_10.
DR   InterPro; IPR001547; Glyco_hydro_5.
DR   InterPro; IPR000119; SLH.
DR   Pfam; PF02018; CBM_4_9; 2.
DR   Pfam; PF00331; Glyco_hydro_10; 1.
DR   Pfam; PF00395; SLH; 2.
DR   PRINTS; PR00134; GLHYDRLASE10.
DR   SMART; SM00633; Glyco_10; 1.
DR   PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR   PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
KW   Xylan degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT   SIGNAL      1     33
FT   CHAIN      34   1157       ENDO-1,4-BETA-XYLANASE A.
FT   DOMAIN    413    649       CATALYTIC (POTENTIAL).
FT   DOMAIN   1055   1113       SLH 1.
FT   DOMAIN   1114   1157       SLH 2.
FT   ACT_SITE  495    495       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE  537    537       NUCLEOPHILE.
FT   ACT_SITE  600    600       D->N: LOSS OF ACTIVITY.
FT   MUTAGEN   537    537       D->N: LOSS OF ACTIVITY.
FT   MUTAGEN   600    600       E->Q: LOSS OF ACTIVITY.
FT   MUTAGEN   602    602       D->N: LOSS OF ACTIVITY.
SQ   SEQUENCE  1157 AA;  128379 MW;  51FA6004497EC58B CRC64;
```

```
Query Match             5.9%;  Score 156; DB 1; Length 1157;
Best Local Similarity   20.4%; Pred. No. 0.019;
Matches   92;  Conservative  64;  Mismatches 128;  Indels 166;  Gaps 23;
```

```
Qy    106  AGLYDPFEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSK 165
           :                  |  ||   :  |:  ||
Db     22  ASLFSP-----PIRVFADDTNINLVSNG---DFESGTIDGWIKQGNPTLAVTTEQAIGQ 72

Qy    166  GNDYVVEAGKTYHFTVQRGP---------------VVTGE-GGNEFAPVQNLQ 209
           :                |||                :|:
Db     73  YSMKVTGRTQTY-----EGPAYSFLGKMQKGESYSVSLKVRLVSGQNSSNPLITVTMFR 126

Qy    210  WSVSGQTV-TLTWQAPASD-------------------KRTYVLNESFD---------- 238
           :|:  |:  ||:|:|||
Db    127  EDDNGKHYDTIVWQKQVSEDSWTTVSGTYTLDYIGTLKTLYMVBSPPDTLEYYIDDVVV 186

Qy    239  -----------------------TQTLPNGWTMIDADGDGHNWLSTIN---VYNTATHTGDGAMFSKS 280
                                  :  |                         |::|
Db    187  TTQNPIQVGNVIANETFENGNT------------------SGWIGTGSSVVKAVYGVA-HSGDYSLLTTG 237

Qy    281  WTAS-GGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANF 339
           |:|       |:  |                  ::|:
Db    238  RTANWNGPSYDL------------TGKIVPGQQYNVDFWVK----------FVNGNDTEQIKA 278

Qy    340  TIKLLEETLGSDKPAPM---------NLVKSEGVKLPAPYQRRTIDLSAYAGQQVLAFRHFN 393
           |  :  :|  :|               :  |  ||:|    ::|:|
Db    279  TVKAT-----SDKDNYIQVNDFANVNKGEWTEIKGSF---TLPVADYSGISIVESQ-N 328

Qy    394  STGIFRLYLDDVAVSGEGSSNDYT--------------YTVYRD---------- 423
                 :|||||||   |||::|          ::|:
Db    329  PT--LEFYIDDFSVIGEISNNQITIQNDIPDLYSVFKDYFPIGVAVDPSRLNDADPHAQL 386

Qy    424  -----NVVIAQNLAATTFNQENVAPGQYNY 448
                |:|:::|
Db    387  TAKHFNMLVAEN---AMKPESLQPTEGNF 412
```

RESULT 6

```
XYND_PAEPO
ID   XYND_PAEPO     STANDARD;       PRT;   635 AA.
AC   P45796;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Endo-1,4-beta-xylanase D precursor (EC 3.2.1.8) (Xylanase D)
DE   (1,4-beta-D-xylan xylanohydrolase D).
```

```
GN   XYND.
OS   Paenibacillus polymyxa (Bacillus polymyxa).
OC   Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX   NCBI_TaxID=1406;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 842;
RX   MEDLINE=92041687; PubMed=1938968;
RA   Gosalbes M.J., Perez-Gonzalez J.A., Gonzalez R., Navarro A.;
RT   "Two beta-glycanase genes are clustered in Bacillus polymyxa:
RT   molecular cloning, expression, and sequence analysis of genes
RT   encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase.";
RL   J. Bacteriol. 173:7705-7710(1991).
CC   -!- FUNCTION: SHOWS XYLANASE ACTIVITY AS WELL AS ALPHA-L-
CC       ARABINOFURANOSIDASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
CC       linkages in xylans.
CC   -!- PATHWAY: Xylan degradation.
CC   -!- SIMILARITY: Belongs to family 43 of glycosyl hydrolases.
CC   -----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------
DR   EMBL; X57094; CAA40378.1; -.
DR   PIR; S19011; S19011.
DR   InterPro; IPR006584; CBD_IV.
DR   InterPro; IPR005084; CBM_6.
DR   InterPro; IPR008979; Gal_bind_like.
DR   Pfam; PF03422; CBM_6; 1.
DR   SMART; SM00606; CBD_IV; 1.
KW   Xylan degradation; Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    635       ENDO-1,4-BETA-XYLANASE D.
SQ   SEQUENCE   635 AA;  67914 MW;  F9DEC6996732316 CRC64;

Query Match                5.2%;   Score 137.5;  DB 1;   Length 635;
Best Local Similarity   21.2%;   Pred. No. 0.14;
Matches   123;  Conservative   65;  Mismatches 176;  Indels 217;  Gaps     32;

Qy   22  TAAAQGGPKTAP------SVTHQAVQCKGIRTSKVKDLRDPIPAGMARIILEAHDVWEDGTGY  77
                           :  :    ||   ||   :   :|     :     |||||:|
Db  156  TADTPIGFWTDPLGKALVTHST-----------------PGMAGVTWLFDPAVLVDDDGTGY 200

Qy   78  QMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKV------------PVNADA 122
                                                           :  ||
Db  201  --LYS------GGGIPNES------DPASIANPKTARVIKLGADMTSVIGSATTIDAPY 245

Qy  123  SFSPT------------------NFVLDGTASADIPAGTYDYVII-NP---------NPG 154
         |  |  :        |          ||  :    :  :: ::  ::   :|||
Db  246  LFEDSGIHKYNGKYYSYCINFA--GTHPQQYPAGEIGTMVSDNPMGPFTYKGHPLKNPY 303

Qy  155  IIYIVG----------------EGYSKGNDYVVRAGKTY---HFT-VQRQGPGDAA 190
         | |  |                :|| | | |         ||  :|| | ||
Db  304  TFFGVGGNNHHAVFNFKNEWYVVVHAQTVSKAQ---IGAGKGYRSPHINKLVHKEDGSIS 360

Qy  191  SVV--VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQTLPNGWTM 248
          :|    |  : ||  |  :|  : | |  : :    :
Db  361  EVQGNMTG-----IAQLSNMNPYTRVEAETIAWQGAGVTTEPT---------------- 397

Qy  249  IDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAK-----------IDLSP 293
                                                           :| : :
Db  398  -QASGGP---ISNLNVTN--IHNGDWIAVGKADFGSAGAKTFKANVATNVGGNIEVRLDS 451

Qy  294  DNYLVTPKVTVPENGKLSYW--VSSQVPWTNEHYGVFLSTTGNEAAN--------FTIKL 343
         :   :  ::   |     :             ||  :  | :
Db  452  ETGPLVGSLKVPSTTGMQTWREVETTINNATGVHNIYLVFTGSGSGNLNLDAWQFTPNT 511

Qy  344  LEETLGSDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHF--NSTGIFRLY 401
```

```
Db  512  GGNTITKVEAENMKIGGTYAGKISAPFD----GVALYANAD-YVSYSQYFANST----- 560
         |  |            | |       :  :         ||     :::: ::::|||
Qy  402  LDDVAVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPK 461
                ::::|||  :            |   :::   ::    |:|      :  ||
Db  561  -HNISVRG-ASSNAGTAKV--DIVIG---------------GVTVGSFNF------TGKTPT 597

Qy  462  VCKDVTVBGSNEFAHVQNLTGSAVGQKVTL------KWDA 495
                      |||         |          :::: :::  |||
Db  598  V------------QTLSNITHATGDQEIKLALTSDDGTWDA 626


RESULT 7
SUBF_BACSU
ID   SUBF_BACSU     STANDARD;     PRT;   1433 AA.
AC   P16397;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Bacillopeptidase F precursor (EC 3.4.21.-) (Esterase) (RP-I protease)
DE   (90 kDa serine proteinase).
GN   BPF OR BPF OR BSU15300.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 195-222.
RX   MEDLINE=90170864; PubMed=2106512;
RA   Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
RA   Pero J.;
RT   "Bacillopeptidase F of Bacillus subtilis: purification of the protein
RT   and cloning of the gene.";
RL   J. Bacteriol. 172:1470-1477(1990).
RN   [2]
RP   REVISIONS.
RC   STRAIN=168;
RX   MEDLINE=90368623; PubMed=2118514;
RA   Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
RA   Pero J.;
RL   J. Bacteriol. 172:5520-5521(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92016713; PubMed=2108961;
RA   Wu X.-C., Nathoo S., Pang A.S.-H., Carne T., Wang S.-L.;
RT   "Cloning, genetic organization, and characterization of a structural
RT   gene encoding bacillopeptidase F from Bacillus subtilis.";
RL   J. Biol. Chem. 265:6845-6850(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
```

RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida H., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [5]
RP   SEQUENCE OF 1-211 FROM N.A.
RX   MEDLINE=89008108; PubMed=3139638;
RA   Beall B., Lowe M., Lutkenhaus J.;
RT   "Cloning and characterization of Bacillus subtilis homologs of
RT   Escherichia coli cell division genes ftsZ and ftsA.";
RL   J. Bacteriol. 170:4855-4864(1988).
RN   [6]
RP   SEQUENCE OF 1410-1433 FROM N.A.
RC   STRAIN=168 / Marburg;
RX   MEDLINE=90174995; PubMed=2106671;
RA   Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;
RT   "Nucleotide sequence of the sporulation gene spoIIGA from Bacillus
RT   subtilis.";
RL   Nucleic Acids Res. 18:657-657(1990).
RN   [7]
RP   SEQUENCE OF 16-219.
RC   STRAIN=Natto 16;
RA   Kato T., Yamagata Y., Arai T., Ichishima E.;
RT   "Purification of a new extracellular 90-kDa serine proteinase with
RT   isoelectric point of 3.9 from Bacillus subtilis (natto) and
RT   elucidation of its distinct mode of action.";
RL   Biosci. Biotechnol. Biochem. 56:1166-1168(1992).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to peptidase family S8.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M29035; AAA62679.1; -.
DR   EMBL; J05400; AAA83362.1; -.
DR   EMBL; Z99111; CAB13403.1; -.
DR   EMBL; M22630; AAA22458.1; -.
DR   EMBL; X17344; CAA35224.1; -.
DR   PIR; A36734; A36734.
DR   HSSP; P00782; 2SBT.
DR   MEROPS; S08.017; -.
DR   SubtiList; BG10233; bpr.
DR   InterPro; IPR008757; Peptidase M6.
DR   InterPro; IPR000209; Peptidase S8.
DR   InterPro; IPR009020; Protease_inhib.
DR   Pfam; PF05547; Peptidase M6; 1.
DR   Pfam; PF00082; Peptidase S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
FT   SIGNAL   1   30
FT   PROPEP   31   194   POTENTIAL.
FT   CHAIN   195   755   BACILLOPEPTIDASE F.
FT   PROPEP   756   1433   POTENTIAL.
FT   ACT_SITE   227   227   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   274   274   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   452   452   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CONFLICT   219   219   T -> A (IN REF. 7).
FT   CONFLICT   393   393   A -> V (IN REF. 3).
FT   CONFLICT   829   834   KHQNKA -> N (IN REF. 3).
FT   CONFLICT   836   841   QPQVLP -> RTRLYS (IN REF. 3).
FT   CONFLICT   844   852   AQVSVVETG -> FCRSRHKSV (IN REF. 3).
FT   CONFLICT   853   1433   MISSING (IN REF. 3).
SQ   SEQUENCE   1433 AA; 154577 MW; 98DF6846897807C9 CRC64;

Query Match   5.1%; Score 135.5; DB 1; Length 1433;
Best Local Similarity   20.6%; Pred. No. 0.54;
Matches 112; Conservative 58; Mismatches 150; Indels 223; Gaps 31;

Qy   90 ASIPEESF---WFANGTIPAGLYDPFEYKV--------PVNADASFSPTNFVLDGTAS----KRTY-   136
           :|:|: :|| | |:| ::|| | :: |::|||:
Db   437 SSVPGQTYEDGW--DGTSMAG---PHVSAVAALLKQANASLSVDEMEDILTSARPLTDS   491

Qy   137 --ADIPAGTYDYVIINPNPGIYIVGEGVSKGNDYVVEAGKT------YHFTVQRQPGD   188
            : ||: : ::: |: | : : || :|
Db   492 TFPDSPNNGYGHGLVNAFDA-VSAVTDGLGKAEGQVSVEGDDQEPPVYQHEKVTEAYEGG   550

Qy   189 AASVVVTGEGNEFAPVQ-----NLQWS------VSGQTVTLTWQAPASD----KRTY-   231
            :|: | |: | | : | :| :| |: ||:
Db   551 SLPLTLTAEDNVSVTSVKLSYKLDQGEWTEITAKRISGDHLKGTYQAEIPDIKGTKLSYK   610

Qy   232 ------------VLNESFDT-----------QTLPNGWTMIDADGDGHRWLSTINVYNT   267
                          :: |:
Db   611 WMIHPFGGHVSSDVYDVTVKPSITAGYKQDFETAPGGWV---ASGTNNNW-----   658

Qy   268 ATHTGDGAMFSKSW-----------TASGGAKI---DLSPDNY----LVTPKVTVPENGK   309
                                    : | |:   : :|: || :| ::
Db   659 ------------EWGVPSTGPNTAASGEKVYGTNLT-GNYANSAMNLVMPPIKAPDSGS   705

Qy   310 LSYWVSSQVPWTNEH----YG-VFLSTTG---NRAANFTKLLEETLGSDKPAPMNLVK   360
           |:: |:: |:| :| || :| ||:: | ||
Db   706 LFLQFKS---WHNLEDDFDYGYVFVLPEGEKNWEQAGVYNGK----------   744

Qy   361 SEGVKLPAPYQERTIDLSAYAGAQQYYLAFRHFNSTGIFR--LYLDDVAVSGEGSSNDYTY   418
           |:| ||||| | ::|||:| | | : |||||:|:|:: | ||::|: : |:|::|::| |
Db   745 ----------TSSWTDEEIDLSAYKGQNIQVMFNIQSDESIAKEGWYIDDVVLSDKSAGK--   794

Qy   419 TVYRDNVVI----------AQNLAATTFNQENVA----   442
           |::|:|:|: |::: | | :|:|
Db   795 TVKKNKLGVEKPSGKQKKKPVNPKKAKPSANTAVKHQNKAIQPQVLPLKAQVSVVETGKS   854

Qy   443 --------PGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVG----QKVTL   491
                   | | | |: : : | || | :|: | : | |:
Db   855 TYSDQSTGQYT---LKHKAG--------DYT--------LMAEAYGYQSKTQKVSL   891

Qy   492 KWD 494
           ||:
Db   892 KTD 894

RESULT 8
OMPB_RICCN
ID   OMPB_RICCN   STANDARD;   PRT;   1655 AA.
AC   Q9KKA3; Q9XK98; Q9XC45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB OR RC1085.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Malish 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";

```
RL   Science 293:2093-2098(2001).
RN   [2]
RP   SEQUENCE OF 33-1649 FROM N.A.
RC   STRAIN=Indian tick typhus, and Malish 7;
RX   MEDLINE=20393643; PubMed=10939649;
RA   Roux V., Raoult D.;
RT   "Phylogenetic analysis of members of the genus Rickettsia using the
RT   gene coding the outer-membrane protein rOmpB (ompB).";
RL   Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN   [3]
RP   SEQUENCE OF 353-1655 FROM N.A.
RC   STRAIN=Malish 7;
RA   Stenos J., Walker D.;
RT   "The rickettsial outer membrane protein A and B genes of Rickettsia
RT   australis, the most divergent rickettsia of the spotted fever group.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC       similarity).
CC   -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC       (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC       layer with hexagonal symmetry (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE008659; AAL03623.1; -.
DR   EMBL; AF123721; AAF34124.1; -.
DR   EMBL; AF123726; AAF34129.1; -.
DR   EMBL; AF149110; AAD39533.1; -.
DR   PIR; E97835; E97835.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   Pfam; PF03797; Autotransporter_1.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 2.
KW   Antigen; S-layer; Cell wall; Complete proteome.
FT   CHAIN         1    1334       120 kDa SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1335    1655       32 kDa BETA PEPTIDE.
FT   VARIANT      61      61       P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT      75      75       G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT      78      78       K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     251     251       N -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     413     413       I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     959     959       A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     988     988       R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT    1139    1139       KD -> GH (IN REF. 3).
FT   CONFLICT    354     354       F -> S (IN REF. 3).
FT   CONFLICT    776     776       E -> D (IN REF. 3).
FT   CONFLICT   1159    1159       G -> S (IN REF. 3).
FT   CONFLICT   1177    1177       H -> R (IN REF. 3).
FT   CONFLICT   1492    1492       E49E19377D5FCE37 CRC64;
SQ   SEQUENCE   1655 AA;  168342 MW;
```

```
Query Match           5.1%;   Score 135;   DB 1;   Length 1655;
Best Local Similarity   19.3%;   Pred. No. 0.7;
Matches 102;  Conservative 53;  Mismatches 189;  Indels 184;  Gaps  20;

Qy    100  ANGTIPAGLYDPFEYKVPVNADASFSPTNFV-------------LDGTASADIPAGTYDYV 147
                |||:         |||: :|  :||          |||||:   :||
Db    439  ANGTLAS------------ASADANVAVTNNITAIEASGAGVVQLSGTHAAELRLG----- 482

Qy    148  IINPNPGIIYIVGEGV--SKGNDYVVEAGKTYHFTVQRQGPGDAASVVTGEGGNE--F 202
                :  :: ||:|   :: :|    :|||     ||||   |||
Db    483  ---NAGSVFKLADGTVINGKVNQTALVGGALAAGTITLDG----SATITGDIGNAGGA 533
```

```
Qy    203  APVQ------------------------------------------------------NLQWSVSGQTVTLTWQ----------------------APASD 227
                |:|                                                    : :||| : :.:|                            ||:|||
Db    534  AALQGITLANDATKTLTLGGANIIGANGTINFQANGGTIKLTSTQNNIVVDFDLAIATD 593

Qy    228  KRTVVLNESFDTQTLPNGWTM-----IDADGGHNWLSTINVYNTATHTGDGAMFSKSWT 282
                |||:|  :|    :|||||          ||:|| |||:|||:||::||| ||:|||:|
Db    594  QTGVV-----DASSLTNAQTLTINGKIGTVGANNKTLGQFNIGSSKTVLSDGDVAINELV 648

Qy    283  ASGGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIK 342
                |||   ||  :|||:|||||||| ||||:||:|||  |||| ||| |||:|||  |||
Db    649  IGNNGAVQFAHNTYLITRTTNAAGQGKIIP------NPVVNNTTL---ATGTNLGSATNP 700

Qy    343  LLEETIGSDKPAPMNLVKS--EGVKLPAPYQERTIDLSAYAGQQVLAFRHFNSTGIFRL 400
                |||:|:|  :| ||:||    |||||   ||||||||||: :| ||:||:||||| |||
Db    701  LAEINFGSKGAANVDTVLNVGKGVNL---YATNITTTDANVGSFIFNA-----GGTNI--- 750

Qy    401  YLDDVAVSGE--GSSNDYTVTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTA-- 456
                 |  ||||||    ||:||:|| |||||:||| ::||||::||:  |||| ||| :|||
Db    751  ------VSGTVGGQQNKFNTVALDNGTTVKFLGNATFNGNTTIAANSTLQIGGNVTADF 804

Qy    457  -----------------------GVSPKVCKDVTVEGS-----NEFAIVQNLTG----- 482
                                        ||:||:|| |||||       |||:||||||
Db    805  VASADGTGIVEFVNTGPITVTLNKQAAPVNALKQITVSGPGNVVINEIGNAGNYHGAVTD 864

Qy    483  ------------------SAVGQKVTLKWDAPN 497
                             ||||||:|||| |||
Db    865  TIAFENSSLGAVVFLPRGIPFNDAGNRIPLTIKSTVGNKTATGFPDVPS 912
```

```
RESULT 9
YA33_SULSO
ID   YA33_SULSO     STANDARD;      PRT;   725 AA.
AC   Q97Z97;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein SSO1033 precursor.
GN   SSO1033.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=11427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC   -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC   -!- SIMILARITY: Contains 6 Kelch repeats.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE006722; AAK41296.1; -.
DR   PIR; A90255; A90255.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR006652; Kelch_rep.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF01344; Kelch; 5.
```

```
DR   SMART; SM00060; FN3; 3.
KW   Hypothetical protein; Kelch repeat; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    725       HYPOTHETICAL PROTEIN SSO1033.
FT   REPEAT       59    100       KELCH 1.
FT   REPEAT      101    145       KELCH 2.
FT   REPEAT      146    199       KELCH 3.
FT   REPEAT      201    248       KELCH 4.
FT   REPEAT      250    297       KELCH 5.
FT   REPEAT      299    342       KELCH 6.
FT   DOMAIN      323    398       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      412    490       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      585    652       FIBRONECTIN TYPE-III 3.
SQ   SEQUENCE    725 AA;    78465 MW;    F8BF1220D013BAB6 CRC64;

Query Match              5.0%;  Score 131.5;  DB 1;  Length 725;
Best Local Similarity   20.4%;  Pred. No. 0.4;
Matches 112; Conservative 61; Mismatches 134; Indels 241; Gaps 27;

Qy  101 NGTIPAGLYDPFEYKVPVNADASFSPTN------------------FVLDGTASADI---- 139
             :|||||   :|
Db  154 NTTSPAGLYFP----PSNAIRLFYPNNDSWRIIGYMPVPTYGGGVVFNGTSLIIVSGYI 208

Qy  140 --PAGTYDYVIINPN----------------------PGIIYIVGEGVSKGNDYVVEAG 174
                                               |:||:
Db  209 GYSAYTNDIILIYSPQNNNWTILNGVLPVWIHDSALAYYRGVLFIVG------GYIYTAG 261

Qy  175 K-------------TYHFTVQRQG----PGDAASVV---------VTGEGNEFAPVQNLQ- 209
        |                 ||||     :  |
Db  262 SGGVNNAILAYYNGNLQRVGLPVPVYSAGYVQVGNMLYLAGGIGSSLSDVSALQLITFN 321

Qy  210 ---------WSVSGQTVTLTWQAPASDKRTYVLNESFDTQTLPNGWTMIDADGDGHNWL 259
                 :   |:|    |||        :|
Db  322 FPPLPPKITSYSAGNESVTLGW----------------NPVRLSSGYEIIYWNNMGFN- 363

Qy  260 STINVYNTATHT----GDGAMF-------------------------------- 277
        :|:|||   :||      | :
Db  364 SSINVGNVTSYTVTGLKDGITYYFEVLAYNSIGYSSPSSIIALTPASVPNPPQLVSVKYG 423

Qy  278 ------SKSW----TASG---GAKIDLSPDNLV-----------------TPKVT---- 303
                         :  : :
Db  424 NDNVTLNWLPTFSGGYLLLGYYIVIVKNENSMVSSHFVNSTSLTISNLTPNVTYNVFIYA 483

Qy  304 VPENGKLSYWVSSQVPWTNEHYGVFLSTTGNE-AANFTIKLLEETIGSDKPAPMNLVKSE 362
        |:|||  |||| |
Db  484 VNKLGNSSPLVLTVPITKASVFAFITKLGNGILVNWTTSF--------PANIT---- 529

Qy  363 GVKLPAPYQERTIDLSAYAGQQVYLAFR----------HPNSTGIFR------LYL- 402
        ::|||                   |||
Db  530 -LELYNPNGNLISQIAAIKGNSSYL-FRVPQGNYTLVIIASNSAGVSKYYQVVVYLPPA 587

Qy  403 ----------DDVAVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEV 452
                                ::|||
Db  588 SPQVSLIGFGNNLYISWNNEANVITTLVYVNNSLVEGPSNSIVT--NISNG--TYLVKV 643

Qy  453 KYTAGVSP 460
        ||:|
Db  644 ---IGVNP 648


RESULT 10
OMPB_RICJA
ID   OMPB_RICJA     STANDARD;      PRT;   1656 AA.
AC   O06653;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB.
OS   Rickettsia japonica.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=35790;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YH;
RA   Uchiyama T.;
RT   "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT   japonica.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC       similarity).
CC   -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC       (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC       layer with hexagonal symmetry.
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   use by non-profit institutions as long as its content is in no way
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AB003681; BAA20138.1; -.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   Pfam; PF03797; Autotransporter;1.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 2.
KW   Antigen; S-layer; Cell wall.
FT   CHAIN         1   1338       120 kDa SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1339   1656       32 kDa BETA PEPTIDE.
FT   DOMAIN      528    533       POLY-GLY.
SQ   SEQUENCE   1656 AA;   168097 MW;   3132A69C9DD5999F CRC64;

Query Match              5.0%;  Score 131;  DB 1;  Length 1656;
Best Local Similarity   21.3%;  Pred. No. 1.3;
Matches  99;  Conservative 45;  Mismatches 198;  Indels 122;  Gaps  18;

Qy  100 ANGTIPAGLYDPFEYKVPVNADASFSPTNFV----------------------LDGTASADIPAGT-- 143
        ||||:||||
Db  439 ANGTLAS----------ASADANVAVTNNITAIRASGVGVVQLSGTHTAELRLGNAGSV 487

Qy  144 ---YDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFTVQRQPG---------------DAA 190
           |:|||
Db  488 FKLADGTVINGKVNQTVLVG-GVLAAGAITLDGSAITTGDIGNGGGAALQSITLANDAT 546

Qy  191 SVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQ--------APASDKRTYVLNESPDTQ--- 240
        |
Db  547 KTITLGGANIISANGGTINFQANGGTIKLTSTQNNIVVDCDLAIATDQTGVV-----DAS 601

Qy  241 TLPNGWTM---------IDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
        ||||||||
Db  602 SLTNAQTLTISGTIGIGANNTLGQFNIGSSKTTLNGGNVAINELVIGNNGSVQFAHNT 661

Qy  296 YILVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPKAP 355
        |||:
Db  662 YLITRTTNAAGQGKIIF----------NPVVNNTTLAAGTNLGSAAN--PLAEINFGS-KGAR 712

Qy  356 MNLVKS--EGVKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDDVAVSGE--G 411
        |||
Db  713 ADTVLNVGEGVNL--YATNITTTDANVGSFV------KNIVSGTVGG 756

Qy  412 SSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTA- 456
        ||
Db  757 QQGNKFNTVALDNGTTVKFLGNATFNGNTTIAANSTLQISGNYTADFIASADGTVEFV 816

Qy  457 --------GVSPKVCKDVTVEGS-----NEFAHVQNLTGS 483
        :|||:   :||  |
```

```
Db    817 NTGPINVTLNKQAVPVNALKQITVSGPGNVVVNEIGNAGNYHGA 860


RESULT 11
OXAA_CHLCV
ID   OXAA_CHLCV     STANDARD;        PRT;    794 AA.
AC   P59809;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Inner membrane protein oxaA.
GN   OXAA OR CCA00472.
OS   Chlamydophila caviae.
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83557;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GPIC;
RX   MEDLINE=22569155; PubMed=12682364;
RA   Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA   Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,
RA   Umayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA   Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,
RA   Fraser C.M.;
RT   "Genome sequence of Chlamydophila caviae (Chlamydia psittaci GPIC):
RT   examining the role of niche-specific genes in the evolution of the
RT   Chlamydiaceae.";
RL   Nucleic Acids Res. 31:2134-2147(2003).
-!- FUNCTION: Required for the insertion of integral membrane proteins
    into the membrane. Probably plays an essential role in the
    integration of proteins of the respiratory chain complexes.
    Involved in integration of membrane proteins that insert
    dependently and independently of the Sec translocase complex (By
    similarity).
-!- SUBUNIT: Specifically interacts with transmembrane segments of
    nascent integral membrane proteins during membrane integration (By
    similarity).
-!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
    (By similarity).
-!- SIMILARITY: Belongs to the OXA1/oxaA family. Subfamily 1.

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).

DR   EMBL; AE016995; AAP05217.1; -.
DR   TIGR; CCA00472; -.
DR   HAMAP; MF_01810; -; 1.
DR   InterPro; IPR001708; 60kDa_innermeb.
DR   Pfam; PF02096; 60KD_IMP; 1.
KW   Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM      5     24       POTENTIAL.
FT   TRANSMEM    571    593       POTENTIAL.
FT   TRANSMEM    639    661       POTENTIAL.
FT   TRANSMEM    691    713       POTENTIAL.
FT   TRANSMEM    740    762       POTENTIAL.
SQ   SEQUENCE   794 AA;  88951 MW;  9EDCF05E73709133 CRC64;


Query Match             4.9%;  Score 129;  DB 1;  Length 794;
Best Local Similarity   21.0%;  Pred. No. 0.65;
Matches   86;  Conservative   63;  Mismatches  154;  Indels  106;  Gaps  21;

Qy     80 LWDADHNQYGASIPEESFW-FANGTIPAGLYDPFEYK----VPVNADASFSPTNFVLDG 133
             ::: ::|      :|:|  :|| |:|:  |:| |: :  :  |:|: :|| :|
Db    167 VYNKDSAVYGTSL---VFWRSGNEYIPLGIYNSKEERLESLDLPITKAAVFSDSKSANAG 223

Qy    134 TASAD--IPAGTYDYVIINPNPGIIYIVGEGV-----SKGNDYVV-EAGKTYHFTVQRQG 185
            :|:||  ||: |::|: ::  ||:| |:|||              :|::| :: |:|:::
Db    224 ANSAQYFVLSNEYMQLIVSQESGSI----EGINLPFSSEDNKSIVNEIG--FDRELKAQV 277
```

```
Qy    186 PGDAASVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASD-
              |:|: ||| |  |||:| :|||   :||||| :|||||
Db    278 PSEASFPGLPSVGANN-QPVSD---TVGGYYPLLRRGIISDAKRT--PSSYHALNIV 329

Qy    244 NGWTMIDADCGHNWLSTINVYNTATHTGDGAM----------------------FSKS 280
              |||  || |||| |:| ||| :| : |||:
Db    330 SGRELTNSVASGYR-VSTFNSTMLELESNDGSIKKTYKLPQQQYAFEVEVGVNRASDDL 388

Qy    281 WTASGGAKIDLSPDNYLVTP---------------------KVTVPE-------NGKLSYWVSSQ 317
             :|:|: ::|| :| : |:|                     ||::|        ||| :|:|||
Db    389 WITSGIPEVEIMSNAF--TPAIKYHVIKKNKGQLDKVKLPKAKDPLALRSGVYPQWILN- 445

Qy    318 VPWINEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVKSEG------------VK 365
             ||:|:|| |   :||  :|: ||:|:::|   :|  ::|:||:
Db    446 ---SNGYFGIILSPLTDIPAGYAASYVP----GSSVPTRLSLLSPKNQAYPASKYPGYETL 499

Qy    366 LPAPYQERTIDLSAYAGQQYYLAFR----HFNSTGIFRLYLDDVAVSG 409
            ||| |::| ||:| :|| |||| |     :||:||| |||:|:||| |
Db    500 LPLPQKEGTHRFLVYAGPLADPTLRVLDKAYTNSKGESPQYLDCITFRG 548


RESULT 12
TOXA_CLODI
ID   TOXA_CLODI     STANDARD;        PRT;    2710 AA.
AC   P16154;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Toxin A.
GN   TOXA OR TCDA.
OS   Clostridium difficile.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1496;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RX   MEDLINE=90221894; PubMed=2103310;
RA   Sauerborn M., von Eichel-Streiber C.;
RT   "Nucleotide sequence of Clostridium difficile toxin A.";
RL   Nucleic Acids Res. 18:1629-1630(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RX   MEDLINE=90129305; PubMed=2105276;
RA   Dove C.H., Wang S.Z., Price S.B., Phelphs C.J., Lyerly D.M.,
RA   Wilkins T.W., Johnson J.L.;
RT   "Molecular characterization of the Clostridium difficile toxin A
RT   gene.";
RL   Infect. Immun. 58:480-488(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RA   von Eichel-Streiber C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
-!- FUNCTION: Only after the enteral delivery of the enterotoxin A may
    the characteristic disease called pseudomembranous colitis be
    induced.
-!- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
    REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
    DIFFERENT OLIGOPEPTIDES.

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).

DR   EMBL; X51797; CAA36094.1; -.
DR   EMBL; M30307; AAA23283.1; -.
```

RL   Gene 147:1-11(1994).
CC   -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC   OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. MAY PLAY A ROLE
CC   IN THE EXPORT OF BUTIROSIN FROM THE ORGANISM.
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC   layer with hexagonal symmetry.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   --
DR   EMBL; L20421; AAA62588.1; -.
DR   PIR; T17884; T17884.
DR   InterPro; IPR003343; Big_2.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF02368; Big_2; 1.
DR   Pfam; PF00395; SLH; 1.
DR   SMART; SM00635; BID_2; 1.
KW   Signal; Cell wall; S-layer.
FT   SIGNAL   1   30   POTENTIAL.
FT   CHAIN   31   1616   S-LAYER RELATED PROTEIN.
SQ   SEQUENCE   1616 AA; 172074 MW; 1F03FF7A4F113AA7 CRC64;

Query Match   4.7%; Score 124; DB 1; Length 1616;
Best Local Similarity 20.1%; Pred. No. 3.5;
Matches 120; Conservative 78; Mismatches 214; Indels 184; Gaps 31;

Qy   1 MRKLNSLFSLAVLLSLL--CWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRD---P   55
              :: : :   ::: : ||::::
Db   5 LRKWNGMMIALVISLLTPAWGKASAQSLGQESGGGV---QPQSAGVTDDVVLSRDGTFR   61

Qy   56 IPAGMARIILEAHDVWEDGTGYQMLWDADHNQYGASI---PEESFWFANGTIPAGLYDPF 112
           |:   : ||
Db   62 LPVGHINVNVDAP---PQKTNVYALFTS------GAQVTNSQETDKVFVKKNTAIAVDKH 113

Qy   113 EYKVPVNADASFSPTNFVLDGTASADIPAGTY----------------------DYV   147
                                                                  |||
Db   114 DQVIRVIGPTAVPPTGSTWEENQNLPIPEGGYVLLANDSSWGTSVTRKPLFEHYKTGDTV 173

Qy   148 --------------------IINPNPGIIYIVGEGVSKGNDYYVEAGKTYHFTVQRQGPGDAASVVV 194
                               :: :   ::
Db   174 SLHKGGQVVHAGDFLNPDPGLNLVTASGTT----VTSPDFTVSGQVVRYGSGQGISLTV 228

Qy   195 TG------------EGGNEFAPVQNLQ--WSVSGQTVTLTW----QAPASDKRTYV   232
              || :   :
Db   229 NGTEAALKADGAFQSAVRLTAGTNAISVKLLKDGREIVSSTVTVTVTNDAQQPA------ 281

Qy   233 LNESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLS 292
                 :   : : :   : ::   :
Db   282 --DLIEVEAAPIDIT-ISIEGPAHA-IGYVD-QDIAGIDDTVALFTNDW---------   325

Qy   293 PDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDK 352
           | |:: ::   :
Db   326 ------------GPQITVPQ------FNVAVQV------------DAGSKVTKVVNPSIDGKT 358

Qy   353 PA---PMNL-VKSEGVKLPAPYQERTIDLSAYAGQQV--YLA----------------   388
           || |:|| :   :   :
Db   359 PAWTGPDLEIPSGGYVLVA--QD-----TSYAGKNIKKYLATYFKVGDAIKLRKNGFAV 411

Qy   389 -FRHFNSTG--IFRLYLDDVAVSGEGS-SNDYTYTVYR----DNVIAQNLAATTENQEN 440
            | |:: | : | : : |:
Db   412 PVKDLMGTGGPIARVTLDNYAMYTETKPSTELSGTITNMDDPSKIALTVNGTPLPFGPDG 471

Qy   441 VAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAH---VQNLTGSAVGQKVTLKW   493
           | ||:|:   : |:|: | |:|
Db   472 KFKTSYTLAEGINYLDIV------VTKEGKEQDSKDLVVYSRPGFSTGKKVIL-W   519

RESULT 14

```
CTA1_BACCI
ID   CTA1_BACCI     STANDARD;      PRT;     972 AA.
AC   P94286;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Cycloisomaltooligosaccharide glucanotransferase precursor (EC 2.4.1.-)
DE   (CITASE).
OS   Bacillus circulans.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1397;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T-3040;
RA   Oguma T., Kurokawa T., Tobe K., Kobayashi M.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PRODUCES CYCLOISOMALTOOLIGOSACCHARIDE FROM DEXTRAN.
CC   -!- SIMILARITY: Belongs to family 66 of glycosyl hydrolases.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D61382; BAA09604.1; -.
DR   InterPro; IPR005084; CBM_6.
DR   InterPro; IPR008979; Gal_bind_like.
DR   Pfam; PF03422; CBM_6; 2.
KW   Transferase; Glycosyltransferase; Signal.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39    972       CYCLOISOMALTOOLIGOSACCHARIDE
FT                                GLUCANOTRANSFERASE.
SQ   SEQUENCE   972 AA;  107431 MW;  2827BEA61782CC22 CRC64;
```

```
Query Match             4.6%;  Score 121.5;  DB 1;  Length 972;
Best Local Similarity  20.8%;  Pred. No. 2.6;  Indels 109;  Gaps 17;
Matches   71;  Conservative  51;  Mismatches  111;

Qy    183 RQGPGDAASVVTGEGGN----------EFAPVQNLQWSVS------GQ--TVTLTWQA 223
              ||||||    |  :               :||:|          :   |     ||
Db     53 RYNPGDAVSIRVQAKNGTGSSWSGAARLEIFHLENSVYTSSQSLSLTNGQSTTLTFTWTA 112

Qy    224 PASDKRTVLNESFDTQTLPNGWTMIDADGD----GHNWLSTINVYNTATHTGDGAMFS 278
              |:::|:|  |:        :|||:  ||:|    |:   ::|:|:     |:|:|
Db    113 PSTDFRGYFVR--IDAGTLGQGATAIDVSSDFFTKYPRYGYISEP------ESGETALES 163

Qy    279 KSWTASGGAKIDLSPDNYLIVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAAN 338
                :|::|            ||   :: :                :|
Db    164 K------AKVDQLAQDY-----------HINAWQFYDWMWRHD---------------- 189

Qy    339 FTIKLLEETLGSDKPAPMNLVKSEGVKLPAPYQERTIDL----------SAYAGQQYYLAFRHF 392
                |::::|:       ::|         |::  :  :            :
Db    190 ---KMIKRTGGSIDSTWLDLFNRE--:ISWSTIQNQIDAVHDVYNGKAMAYAMIYASRENY 243

Qy    393 NSTGI--FRLYLDDVAVS-----GEGSSNDYTYTVRDNVVIAQNLAATTFNQENVAP 443
              ||::   :::  ||          |:   | :||
Db    244 SPLGISPTWGIYEDSSHTNQFDVDFGDGS----TYLYMSD----PQN----------P 283

Qy    444 GQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAV 485
              ::|  |  : ::::  :|
Db    284 NWQNY-IHAEYIDSINTAGFDGIHVDQMGQRSNVYDYNGNSI 324
```

```
RESULT 15
PICP_PSESR
ID   PICP_PSESR     STANDARD;      PRT;     587 AA.
AC   P42790;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
```

```
DE   Pseudomonalisin precursor (EC 3.4.21.100) (Pepstatin-insensitive
DE   carboxyl proteinase) (Pseudomonapepsin).
GN   PCP.
OS   Pseudomonas sp. (strain 101) (Achromobacter parvulus T1).
OC   Bacteria; Proteobacteria.
OX   NCBI_TaxID=33067;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 216-224.
RX   MEDLINE=95014496; PubMed=7929375;
RA   Oda K., Takahashi T., Tokuda Y., Shibano Y., Takahashi S.;
RT   "Cloning, nucleotide sequence, and expression of an isovaleryl
RT   pepstatin-insensitive carboxyl proteinase gene from Pseudomonas sp.
RT   101.";
RL   J. Biol. Chem. 269:26518-26524(1994).
RN   [2]
RP   SEQUENCE OF 216-585.
RX   MEDLINE=96157710; PubMed=8576087;
RA   Hayashi K., Izu H., Oda K., Fukuhara K.-I., Matsuo M., Takano R.,
RA   Hara S.;
RT   "The primary structure of pepstatin-insensitive carboxyl proteinase
RT   produced by Pseudomonas sp No. 101.";
RL   J. Biochem. 118:738-744(1995).
RN   [3]
RP   MUTAGENESIS OF ASPARTIC ACID AND GLUTAMIC ACID RESIDUES, AND ACTIVE
RP   SITES.
RX   MEDLINE=99419069; PubMed=10488127;
RA   Oyama H., Abe S.-I., Ushiyama S., Takahashi S., Oda K.;
RT   "Identification of catalytic residues of pepstatin-insensitive
RT   carboxyl proteinases from prokaryotes by site-directed mutagenesis.";
RL   J. Biol. Chem. 274:27815-27822(1999).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 216-585.
RX   MEDLINE=21614391; PubMed=11747435;
RA   Wlodawer A., Li M., Gustchina A., Dauter Z., Uchida K., Oyama H.,
RA   Goldfarb N.E., Dunn B.M., Oda K.;
RT   "Inhibitor complexes of the Pseudomonas serine-carboxyl proteinase.";
RL   Biochemistry 40:15602-15611(2001).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.0 ANGSTROMS) OF 216-587.
RX   MEDLINE=21223584; PubMed=11323721;
RA   Wlodawer A., Li M., Dauter Z., Gustchina A., Uchida K., Oyama H.,
RA   Dunn B.M., Oda K.;
RT   "Carboxyl proteinase from Pseudomonas defines a novel family of
RT   subtilisin-like enzymes.";
RL   Nat. Struct. Biol. 8:442-446(2001).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of the B chain of insulin at 13-
CC       Glu-|-Ala-14, 15-Leu-|-Tyr-16 and 25-Phe-|-Tyr-26 and angiotensin
CC       I at 4-Tyr-|-Ile-5. A good synthetic substrate is Lys-Pro-Ile-Glu-
CC       Phe-|-Phe(NO(2))-Arg-Leu.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit.
CC   -!- SUBCELLULAR LOCATION: Periplasmic.
CC   -!- PTM: Autocatalytically processed.
CC   -!- SIMILARITY: Belongs to peptidase family S53.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D37970; BAA07188.1; -.
DR   PDB; 1GA1; 11-FEB-03.
DR   PDB; 1GA4; 11-FEB-03.
DR   PDB; 1GA6; 11-FEB-03.
DR   PDB; 1KDV; 11-FEB-03.
DR   PDB; 1KDY; 11-FEB-03.
DR   PDB; 1KD2; 05-FEB-03.
DR   PDB; 1KE1; 11-FEB-03.
DR   PDB; 1KE2; 11-FEB-03.
DR   MEROPS; S53.001; -.
DR   Pfam; PF00082; Peptidase_S8; 1.
```

```
KW   Hydrolase; Protease; Serine protease; Zymogen; Periplasmic; Signal;
KW   3D-structure; Calcium-binding;
FT   SIGNAL        1      ?      POTENTIAL.
FT   PROPEP        ?      215    REMOVED IN MATURE FORM.
FT   CHAIN         216    585    PSEUDOMONALISIN.
FT   PROPEP        586    587    REMOVED IN MATURE FORM.
FT   ACT_SITE      295    295    CHARGE RELAY SYSTEM.
FT   ACT_SITE      299    299    CHARGE RELAY SYSTEM.
FT   ACT_SITE      502    502    CHARGE RELAY SYSTEM.
FT   DISULFID      352    391
FT   METAL         543    543    CALCIUM.
FT   METAL         544    544    CALCIUM (VIA CARBONYL OXYGEN).
FT   METAL         559    559    CALCIUM (VIA CARBONYL OXYGEN).
FT   METAL         561    561    CALCIUM (VIA CARBONYL OXYGEN).
FT   METAL         563    563    CALCIUM.
SQ   SEQUENCE      587 AA;  61072 MW;  E193D8B2C225829A CRC64;

Query Match          4.6%;  Score 120.5;  DB 1;  Length 587;
Best Local Similarity   23.5%;  Pred. No. 1.6;
Matches   78;  Conservative   33;  Mismatches  108;  Indels  113;  Gaps   16;

Qy    103 TIPAGL---YDPFEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPFGIIYIV 159
              ||:|       ::|  |:  ||| :|| :|               ||||||
Db    214 TLAAGTAKGHNPTEF--PTIYDASSAPT----------------------AANTTVGIITIG 251

Qy    160 G--------EGVSKGNDYVVEAGKT---YHFTVQRQGPGDAASVVVTGEGGNEF 202
                      |:|       :: |:     :      :|
Db    252 GVSQTIQDLQQFTSANGLASVNTQTIQTGSSNGDYSDDQQGQGEWDLDSQSIVGSAG--- 308

Qy    203 APVQNLQW-----SVSGQT-VTLTWQAPASDKRTYVLNESFDTQTLPNGWTMIDADGDGH 256
              ||||         ||       :  |:              ||:
Db    309 GAVQQILFYMADQSASGNTGLTQAFNQAVSDNVAKVINVSL------GWCEADANADG- 360

Qy    257 NWLSTINVYNTATHTG----------------DGAMFSKSWTASGGAKIDLSPDN 295
              :  ::  ||                   ||::|  :|
Db    361 TLQAEDRIFATAAAQGQTFSVSSGDEGVYECNNRGYPDGSTYSVSWPAS------SPNV 413

Qy    296 YLVTPKVTVPENGKLSYWVSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLG----SD 351
              |:  ::||:|:|       |:|:|:|             ||  ||
Db    414 IAV-------GGTTLYTTSAGAYSNE------TVWNEGLDSNGKLWATGGGYSVYES 457

Qy    352 KPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQ 383
              ||:  ::|       |:
Db    458 KPSWQSVVSG------TPGRRLLPDISFDAAQ 483
```

Search completed: May 18, 2004, 11:33:18
Job time : 8.24316 secs

```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        May 18, 2004, 11:28:29 ; Search time 19.8443 Seconds
                                        (without alignments)
                             7902.136 Million cell updates/sec

Title:          US-08-353-485-2
Perfect score:  2641
Sequence:       1 MRKLNSLFSLAVLLSLLCWG.........QNLTGSAVGQKVTLKWDAPN 497
Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5
Searched:       1017041 seqs, 31551820 residues

Total number of hits satisfying chosen parameters:          1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  SPTREMBL_25:*
             1: sp_archea:*
             2: sp_bacteria:*
             3: sp_fungi:*
             4: sp_human:*
             5: sp_invertebrate:*
             6: sp_mammal:*
             7: sp_mhc:*
             8: sp_organelle:*
             9: sp_phage:*
            10: sp_plant:*
            11: sp_rodent:*
            12: sp_virus:*
            13: sp_vertebrate:*
            14: sp_unclassified:*
            15: sp_rvirus:*
            16: sp_bacteriap:*
            17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	951.5	36.0	1706	2	Q51839	Q51839 porphyromon
2	950.5	36.0	1706	2	Q51838	Q51838 porphyromon
3	944.5	35.8	1704	2	Q51816	Q51816 porphyromon
4	943.5	35.7	1687	2	Q9R9B7	Q9r9b7 porphyromon
5	824	31.2	1723	2	P72194	P72194 porphyromon
6	818	31.0	1723	2	P72197	P72197 porphyromon
7	816	30.9	1358	2	P96967	P96967 porphyromon
8	812.5	30.8	1732	2	Q51817	Q51817 porphyromon
9	807.5	30.6	1732	2	O52050	O52050 porphyromon
10	806.5	30.5	1732	2	O07442	O07442 porphyromon
11	805.5	30.5	1223	2	Q9ZNB5	Q9znb5 porphyromon
12	561.5	21.3	1097	2	P72196	P72196 porphyromon
13	379	14.4	312	2	Q9KIB3	Q9kib3 porphyromon
14	337	12.8	925	2	Q9F4J0	Q9f4j0 porphyromon
15	213.5	8.1	293	2	Q9XBU9	Q9xbu9 porphyromon
16	156	5.9	1454	16	Q8ENZ2	Q8enz2 oceanobacil
17	154	5.8	1155	17	Q8TPZ8	Q8tpz8 methanosarc
18	153	5.8	5291	16	Q8X2T1	Q8x2t1 escherichia
19	152	5.7	1234	2	Q60046	Q60046 thermoanaer
20	150	5.6	1483	16	Q977Y4	Q977y4 clostridium
21	147	5.6	1983	16	Q9CJ55	Q9cj55 lactococcus
22	146	5.5	2951	17	Q8THC9	Q8thc9 methanosarc
23	144.5	5.5	781	16	Q9S2G1	Q9s2g1 streptomyce
24	139	5.3	5020	16	Q8E9W3	Q8e9w3 shewanella
25	138.5	5.2	2468	16	Q9I2M3	Q9i2m3 pseudomonas
26	137.5	5.2	8173	16	Q7UDU8	Q7udu8 rhodopirell
27	137	5.2	2013	16	Q92EK2	Q92ek2 listeria in
28	135	5.1	1615	2	Q9F0P9	Q9f0p9 rickettsia
29	135	5.1	1654	2	Q93QW9	Q93qw9 rickettsia
30	135	5.1	3064	16	Q82XT8	Q82xt8 nitrosomona
31	133	5.0	897	16	Q8G7K7	Q8g7k7 bifidobacte
32	133	5.0	1734	17	Q8TI67	Q8ti67 methanosarc
33	132	5.0	698	16	Q82F44	Q82f44 streptomyce
34	131.5	5.0	706	16	Q93J50	Q93j50 streptomyce
35	131.5	5.0	768	16	Q88G76	Q88g76 pseudomonas
36	131.5	5.0	1379	17	Q9HL55	Q9hl55 thermoplasm
37	131.5	5.0	1618	2	Q9KKB4	Q9kkb4 rickettsia
38	131	5.0	1357	17	Q8THC8	Q8thc8 methanosarc
39	131	5.0	1618	2	Q9KKB1	Q9kkb1 rickettsia
40	131	5.0	2009	16	Q7UY44	Q7uy44 rhodopirell
41	131	5.0	2768	16	Q8E9G6	Q8e9g6 shewanella
42	130.5	4.9	524	16	Q7UUW6	Q7uuw6 rhodopirell
43	130.5	4.9	1616	2	Q9KKB0	Q9kkb0 rickettsia
44	130.5	4.9	1617	2	Q9KKB9	Q9kkb9 rickettsia
45	130.5	4.9	2217	17	Q8TP72	Q8tp72 methanosarc

ALIGNMENTS

```
RESULT 1
Q51839
ID   Q51839        PRELIMINARY;      PRT;   1706 AA.
AC   Q51839; Q51840;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arginine-specific thiol protease precursor.
GN   PRTR.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96160709; PubMed=7857299;
RA   Kirszbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N.,
RA   Reynolds E.C.;
RT   "Complete nucleotide sequence of a gene prtR of Porphyromonas
RT   gingivalis W50 encoding a 132 kDa protein that contains an arginine-
RT   specific thiol endopeptidase domain and a haemagglutinin domain.";
RL   Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96311339; PubMed=8713096;
RA   Slakeski N., Cleal S.M., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtR that encodes
RT   an arginine-specific thiol proteinase and multiple adhesins.";
RL   Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Reynolds E.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
```

```
RA   Slakeski N.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L26341; AAC18876.1; -.
HSSP; P95493; 1CVR.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN       228   1706       ARGININE-SPECIFIC THIOL PROTEASE.
SQ   SEQUENCE   1706 AA;  185626 MW;  E8BDF07C9813B844 CRC64;

Query Match             36.0%;  Score 951.5;  DB 2;  Length 1706;
Best Local Similarity   31.9%;  Pred. No. 2.7e-50;
Matches  237;  Conservative  74;  Mismatches  156;  Indels  275;  Gaps  20;

Qy     22 TAAAQG--------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
              ||||        |:|  ::|   ||   |||      :::||    |||:||:||||||
Db    671 TATTGGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
           |  :||:|:||||   ||  ::|:||   :||||||||||||  :|||||||||||||||
Db    731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy    127 TNFVLDGTASADIPAGTYDVVIINPNGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
           ||:|  ::|:   |||||||:|  :|||  :||:||||:|   :|  ||||:||:|||
Db    790 TNIMDGIASVNIFAGTYDFAIAAPQANAKIWIAGQGPTKEDDVVFEAGKKYHFLMKKMG 849

Qy    186 PGDAASVV---------------------------- 193
              : :
Db    850 SGDGTELTISEGGGSDYTTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909

Qy    194 -------VTGEGGNEFAPVQNLQWSVSGQTVLTWQAPAS------------DKRTYVLNE 235
                   ::|  ||  |:  |||  :: ||:|    |||            |:|||||:|
Db    910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTTLSE 969

Qy    236 SFDTCTLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
           |  |  |:|| || ||||||||   :||:|  :|:|:||  ||:||||:|||  |:|||
Db    970 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1025

Qy    296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETL---GSD 351
           ||:  |:  |::|||   :|:: :||||||:|   |:||||:|| :|||||||    |||
Db   1026 YLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1085

Qy    352 KPAPM----------------------------------- 356
            ||||
Db   1086 SPEAMRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1145

Qy    357 -----------------------------NLVKS------------------------ 361
                                         :|:|
Db   1146 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALNPDNYL 1205

Qy    362 ------------------------------------------------------------ 361
Db   1206 ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1265

Qy    362 ---EGVKLFAPYQERTIDLSAYAGQVYILAFRHFNSTGIFRLYLDDV--AVSGEGSSND 415
               :||:||  ||   :|:|:|:||  |||||||:|||| ||||||  ||:|| ::||
Db   1266 STEADGAKPQQSVWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD 1322

Qy    416 YTYTVVRDNVVIAQNLAATTPNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA 475
           |  ||||:|  || | :    :||||||||:|||||||||||||||||:| |:||||||
```

```
Db   1323 YTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN 1381
Qy    476 HVQNLTGSAVGQKVTLKWDAPN 497
           |:|||||||||||:|||:|:||
Db   1382 PVKNLKAQPDGGDVVLKWEAPS 1403
```

```
RESULT 2
Q51838
ID   Q51838           PRELIMINARY;   PRT;   1706 AA.
AC   Q51838;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Protease precursor.
GN   PRPR1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96071894; PubMed=7591131;
RA   Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.;
RT   "Characterization, genetic analysis, and expression of a protease
RT   antigen (PrpRI) of Porphyromonas gingivalis W50.";
RL   Infect. Immun. 63:4744-4754(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.;
RT   "The prpR1 and the prR2 arginine-specific protease genes of
RT   Porphyromonas gingivalis W50 produce five biochemically distinct
RT   enzymes.";
RL   Mol. Microbiol. 23:0-0(1997).
DR   EMBL; X82680; CAA57997.1; -.
HSSP; P95493; 1CVR.
MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN       228    719       ALPHA-PROTEASE.
FT   CHAIN       720   1262       BETA-ADHESIN.
SQ   SEQUENCE   1706 AA;  185705 MW;  0E56DCD87FDA8CDD CRC64;

Query Match             36.0%;  Score 950.5;  DB 2;  Length 1706;
Best Local Similarity   31.8%;  Pred. No. 3.1e-50;
Matches  236;  Conservative  77;  Mismatches  154;  Indels  275;  Gaps  20;

Qy     22 TAAAQG--------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
              ||||        |:|  ::|   ||   |||      :::||    |||:||:||||||
Db    671 TATTGGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
           |  :||:|:||||   ||  ::|:||   :||||||||||||  :|||||||||||||||
Db    731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy    127 TNFVLDGTASADIPAGTYDVVIINPNGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
           ||:|  ::|:   |||||||:|  :|||  :||:||||:|   :|  ||||:||:|||
```

```
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007111; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   Pfam; PF01364; Peptidase_C25_C.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1, 1.
FT   CHAIN        228    719       MATURE 50-KDA CYSTEINE PROTEINASE
FT                                 GINGIPAIN.
SQ   SEQUENCE   1704 AA;  185436 MW;  6A34B40131C2A676 CRC64;

Query Match             35.8%;   Score 944.5;  DB 2;  Length 1704;
Best Local Similarity   31.9%;   Pred. No. 7.4e-50;
Matches  235;  Conservative   75;  Mismatches  160;  Indels  267;  Gaps  20;

Qy     22 ----------GPKTAPSVTHQAVQK--GIR--TSKVKDLRDPIPAGMARIILEAH 68
                      : ||| ::|     |   |:|:||       |:||
Db    671 TATTQGQKVTLKWDAPSTKNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
             :|: ||: :| :  ::|:: |::|||   :|  :|:||||::||||::||||
Db    731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTIM-PNCSVPANLFAPFEYTVPENADDPSCSP 789

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
          ||  | |||: ||||||||:::: |||   |:|      |||||||:|||:|| |::|
Db    790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy    186 PGDAASVV- 193
          ||||||||
Db    850 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909

Qy    194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTYVLNESF 237
                :||:||||||||:| ::|||:|||| ||:|          |:||:|:|
Db    910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESF 969

Qy    238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
          : |  |||| |||||||| |     | || :|   :   : |  :|:  | || |||||
Db    970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy    298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS------ 350
          :|| : :|:|||| :||| | :: ||||||:  ||||:  ||| :||||||
Db   1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy    351 -------------KSEGVKL- 366
                       |:|| :|
Db   1086 EAIRGRIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145

Qy    367 ---PAPYQERTID--------LSAYAGQQV- 385
             || :    ||        ||: :||:
Db   1146 STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy    386 --------------------- 385
Db   1206 KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST 1265

Qy    386 ---------------YILAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
                         |:|||||::| ::| |:||||   |:|| ||::||||||
Db   1266 EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTV 1325

Qy    421 YRDNVVIAQNLAATFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
          |||  :||: | :| |:::|| |:|:|||||||||||||:| ::||::  ::| |:|:
Db   1326 YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1384

Qy    481 TGSAVGQKVTLKWDAPN 497
          | |  || ||||| || 
Db   1385 KAQPDGGDVVLKWEAPS 1401
```

```
Db    790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849
Qy    186 PGDAASVV- 193
Db    850 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVS 909
Qy    194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTYVLNE 235
Db    910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSE 969
Qy    236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
Db    970 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1025
Qy    296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
Db   1026 YLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1085
Qy    351 --------------DKPA----- 354
Db   1086 SPEAIRGRIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1145
Qy    355 --------------PMNLV 359
Db   1146 ESSTHGEATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALNPDNYL 1205
Qy    360 ----------------- 359
Db   1206 ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1265
Qy    360 --KSEGVKLPAPYQERTIDLSAYAGQQVLAFRHFNSTGIFRLYLDDV--AVSGEGSSND 415
Db   1266 STEADGAKPQSVWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD 1322
Qy    416 YTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA 475
Db   1323 YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN 1381
Qy    476 HVQNLTGSAVGQKVTLKWDAPN 497
Db   1382 PVRNLKAQPDGGDVVLKWEAPS 1403
```

```
AC    Q9R9B7;
DT    01-MAY-2000 (TrEMBLrel. 13, Created)
DT    01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE    Hemagglutinin/protease.
GN    HAGE.
OS    Porphyromonas gingivalis (Bacteroides gingivalis).
OC    Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC    Porphyromonadaceae; Porphyromonas.
OX    NCBI_TaxID=837; -.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=381;
RA    Han N., Dong H., Progulske-Fox A.;
RT    "Cloning and characterization of hagE from P. gingivalis 381.";
RL    Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AF026946; AAD01810.1; -.
HSSP; P95493; 1CVR.
MEROPS; C25.001.;
DR    GO; GO:0005524; F:ATP binding; IEA.
DR    GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR    GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR    GO; GO:0006310; P:DNA recombination; IEA.
DR    GO; GO:0006281; P:DNA repair; IEA.
DR    GO; GO:0006260; P:DNA replication; IEA.
DR    GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR    InterPro; IPR000977; DNA_ligase.
DR    InterPro; IPR007110; Ig-like.
DR    InterPro; IPR001769; Peptidase_C25.
DR    InterPro; IPR005536; Peptidase_C25_C.
DR    Pfam; PF01364; Peptidase_C25; 1.
DR    Pfam; PF05536; Peptidase C25 C; 1.
DR    Pfam; PF03785; Peptidase C25 C; 1.
DR    PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW    Protease.
SQ    SEQUENCE  1687 AA;  183702 MW;  D085B516A399FE70 CRC64;

Query Match      35.7%;    Score 943.5;   DB 2;   Length 1687;
Best Local Similarity  31.8%;   Pred. No. 8.4e-50;
Matches 235;   Conservative 75;   Mismatches 160;   Indels 269;   Gaps 20;

Qy     22  TAAAQG---------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH   68
                                                                |:|||::
Db    652  TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH   711

Qy     69  DVWEDGTGYQMLWDADHNQYGASIPEE---SFWFANGTIPAGLYDPFEYKVPVNADASFSP   126
Db    712  DVWNDGSGYQLLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP    770

Qy    127  TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG    185
Db    771  TNMIMDGTASVNIPAGTYDYFAIAAPQANAKIWIAGQGPTKEDDVVFEAGKKYHFLMKKMG   830

Qy    186  PGDAASVV----------------------   193
Db    831  SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS   890

Qy    194  -----VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS-----------DKRTYVLNE    235
Db    891  PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTTLSE    950

Qy    236  SFDTQTLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDN    295
Db    951  SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN   1006

Qy    296  YLVTPKVTVPBNGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFIKLLBETLGS----    350
Db   1007  YLITPALDLPNGGKLIFWVCAQDANYASEHYAVYASSTGNDASNFTNALLETITAKGVR   1066

Qy    351  -------------DKPAPMNLV-----------KSEGVKL------    366
Db   1067  SPEAIRGRIGQTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF   1126

Qy    367  ------PAPYQERTID---------------------LSAYAGQV-------    385
Db   1127  ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL   1186

Qy    386  -------------------------------------    385
Db   1187  ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL   1246

Qy    386  -------------------------------------    386
Db   1247  STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYTY   1306

Qy    419  TVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKTAGVSPKVCKDVTVEGSNEFAHVQ   478
           |:|:| |                           |
Db   1307  TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVK  1365

Qy    479  NLTGSAVGQKVTLKWDAPN   497
           ||   | | :
Db   1366  NLKAQPDGGDVVLKWEAPS   1384


RESULT 5
P72194
ID    P72194       PRELIMINARY;      PRT;    1723 AA.
AC    P72194;
DT    01-FEB-1997 (TrEMBLrel. 02, Created)
DT    01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE    Lys-gingipain.
GN    KGP.
OS    Porphyromonas gingivalis (Bacteroides gingivalis).
OC    Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC    Porphyromonadaceae; Porphyromonas.
OX    NCBI_TaxID=837; -.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=381;
RX    MEDLINE=97044756; PubMed=8889827;
RA    Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
RT    "Cloning and sequencing of the gene encoding a novel lysine-specific
RT    cysteine proteinase (Lys-gingipain) in Porphyromonas gingivalis:
RT    structural relationship with the arginine-specific cysteine proteinase
RT    (Arg-gingipain).";
RL    J. Biochem. 120:398-406(1996).
DR    EMBL; D83258; BAA11870.1; -.
DR    MEROPS; C25.002;.
DR    GO; GO:0005524; F:ATP binding; IEA.
DR    GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR    GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR    GO; GO:0006310; P:DNA recombination; IEA.
DR    GO; GO:0006281; P:DNA repair; IEA.
DR    GO; GO:0006260; P:DNA replication; IEA.
DR    GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR    InterPro; IPR000977; DNA_ligase.
DR    InterPro; IPR001769; Peptidase_C25.
DR    InterPro; IPR005536; Peptidase_C25_C.
DR    Pfam; PF01364; Peptidase_C25; 1.
DR    Pfam; PF03785; Peptidase_C25_C; 1.
DR    PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ    SEQUENCE  1723 AA;  187261 MW;  5628963D251493EB CRC64;

Query Match       31.2%;    Score 824;   DB 2;   Length 1723;
Best Local Similarity  29.8%;   Pred. No. 2.3e-42;
Matches 220;   Conservative 75;   Mismatches 170;   Indels 274;   Gaps 21;

Qy     22  TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD-------------PIPAGMARIILEA   67
Db    693  TATTQGQKVTLKWDAPS----AKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAA   747

Qy     68  HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP   126
Db    748  DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVTT   806

Qy    127  TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV   181
```

```
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1723 AA;  186831 MW;  4508A7E50197CEBD CRC64;

Query Match     31.0%;  Score 818;  DB 2;  Length 1723;
Best Local Similarity   29.4%;  Pred. No. 5.4e-42;
Matches  218;  Conservative  78;  Mismatches 166;  Indels 280;  Gaps  21;

Qy     22 TAAAQGGPKI-----APSVTHQAVQKGIRTSKVKDLRD----------PIPAGMARIILEA 67
Db    693 TATTQGQKVTLKWDAPS----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy     68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
Db    748 DNVWGDNTGYQFLLDADHNTPGSVIPATCPLF-TGTASSNLYSANFEYLIPANADPVVTT 806

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
Db    807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 866

Qy    182 QRQGPGDAASVV- 193
Db    867 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy    194 -------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------DKRTYVLNE 235
Db    927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSE 986

Qy    236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
Db    987 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1042

Qy    296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
Db   1043 YLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1102

Qy    351 --DKPA 354
Db   1103 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1162

Qy    355 -----PMNLV 359
Db   1163 ESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL 1222

Qy    360 --- 359
Db   1223 ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1282

Qy    360 --KSEGVKLPAPYQERTIDLSAYAGQQVLAFRHFNSTGIFRLYLDDV--AVSGEGSSND 415
Db   1283 STEADGAKPQSVWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTTD 1339

Qy    416 YTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA 475
Db   1340 YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTI-NPTQFN 1398

Qy    476 HVQNLTGSAVGQKVTLKWDAPN 497
Db   1399 PVKNLKAQPDGGDVVLKWEAPS 1420
```

RESULT 7

```
P96967
ID   P96967       PRELIMINARY;        PRT;   1358 AA.
AC   P96967;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin.
```

RESULT 6

```
P72197
ID   P72197       PRELIMINARY;        PRT;   1723 AA.
AC   P72197;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA   Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT   "Molecular cloning and characterization of Porphyromonas gingivalis
RT   Lys-gingipain.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U54691; AAA99810.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
```

```
GN   HAGD.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Lepine G., Whitlock J., Wojciechowski L., Progulske-Fox A.;
RT   "Cloning, sequencing and characterization of hagD, a member of the
RT   HArep multigene family in Porphyromonas gingivalis.";
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U68468; AAB49691.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   SMART; SM00060; FN3; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1358 AA;  147102 MW;  47FCA0B25B06DED8 CRC64;

Query Match             30.9%;  Score 816;  DB 2;  Length 1358;
Best Local Similarity   29.6%;  Pred. No. 5.1e-42;
Matches  219;  Conservative   75;  Mismatches  171;  Indels  274;  Gaps   21;

Qy     22 TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD---------PIPAGMARIILEA 67
             :   :||  |||   :     ::|||||                    |:|||||||||
Db    328 TATTQGQKVTLKWDAPS----------AKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAA 382

Qy     68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
           ||| |:||:||: | ||||:  |||||   | |  ||| ||:|| ||| ||||||  ::
Db    383 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 441

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
            :::: | : ::||.|||||| | ||:|  | : : .:    ::. ||  : ||| : |:
Db    442 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGDNQPARYDDFTFEAGKKYTFTM 501

Qy    182 QRQGPGDAASVV------ 193
           :|:| |:|   :|
Db    502 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 561

Qy    194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS--------DKRTYVLNE 235
                   :||| |||||||||| .|.| :| ||: | |.:        ||:|:||:|
Db    562 AGVSPKVCVKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNPNPNPNPNPGTTTLSE 621

Qy    236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
           || :  :|  | ||||||| :  |   |  :|  | :||   |  ||.   | :||:|||
Db    622 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 677

Qy    296 YLVIPKVTVPENGKLSYWVSSQ-VEWTNEHYGVFLSTTGNEAANFTIKLLEETLGS--- 350
           ||:||  :  :|||: :||  | : :::|||.:  :::||: ::|| ::||||:    
Db    678 YLITPALDLANGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEETITAKGVR 737

Qy    351 ----------DKPAPMNLV----------------KSEGVKL- 366
                     |:|   ::|                :||   |
Db    738 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 797

Qy    367 ----------PAPYQERTID--------------LSAYAGQQV--- 385
                    |:| ::: :|              | :|.| |
Db    798 ESSTHGEAPAEWTTIDADGDGQDWLCLSSQLDWLTAHGGTNVVASFSWNGMALNPDNYL 857

Qy    386 QRQGPGDAASVV-- 385
           :|:|
Db    858 ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 917
```

```
Qy    386 -----------------------YLAFRHFNSTGIFRLJYLDDV--AVSGEGSSNDYTY 418
                                  :||||:|:| ||:|||::||||  ||:|| ::||||
Db    918 STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTY 977

Qy    419 TVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCVNVTI-NPTQFNPVK 478
           ||||| |.|:| | ||||::|.|| |:::||||||||||||||||||||| |||||||||
Db    978 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNPVK 1036

Qy    479 NLTGSAVGQKVTLKWDAPN 497
           ||: :  | ||:||| ||.
Db   1037 NLKAQPDGGDVVLKWEAPS 1055
```

```
RESULT 8
Q51817

ID   Q51817      PRELIMINARY;      PRT;   1732 AA.
AC   Q51817;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Porphypain.
GN   PRTP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RC   SEQUENCE FROM N.A.
RC   STRAIN=W12;
RX   MEDLINE=96213011; PubMed=8631659;
RA   Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA   Progulske-Fox A., Lantz M.S.;
RT   "Analysis of the prtP gene encoding porphypain, a cysteine proteinase
RT   of Porphyromonas gingivalis.";
RL   J. Bacteriol. 178:2734-2741(1996).
DR   EMBL; U42210; AAB06565.1; -.
DR   PIR; T30836; T30836.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1732 AA;  187875 MW;  654271DBEF7BCAE4 CRC64;

Query Match             30.8%;  Score 812.5;  DB 2;  Length 1732;
Best Local Similarity   30.0%;  Pred. No. 1.2e-41;
Matches  222;  Conservative   70;  Mismatches  172;  Indels  277;  Gaps   22;

Qy     22 TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD---------PIPAGMARIILEA 67
             :   :||  |||   :     ::|||||                    |:|||||||||
Db    693 TATTQGQKVTLKWEAPS----------AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy     68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
           ||| |:||:||: | ||||:  |||||   | |  ||| ||:|| ||| ||||||  ::
Db    748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVTT 806

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
            :::: | : ::||.|||||| | ||:|  | : : .:    ::. ||  : ||| : |:
Db    807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 866

Qy    182 QRQGPGDAASVV------ 193
           :|:| |:|   :|
Db    867 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926
```

```
KW   Protease.
SQ   SEQUENCE   1732 AA;   187931 MW;   B2337463D5CB5EA5 CRC64;

     Query Match              30.6%;  Score 807.5;  DB 2;  Length 1732;
     Best Local Similarity    30.0%;  Pred. No. 2.4e-41;
     Matches  222;  Conservative  69;  Mismatches 173;  Indels 277;  Gaps  22;

Qy     22  TAAAQGGPKI----APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARIILEA    67
               :  :|| :|:|                            |:::|
Db    693  TATTGQKVTLKWEAPS-----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVLAA   747

Qy     68  HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-PEYKVPVNADASFSP  126
            ||| | ||||| ||||||| | |  ||      ||| | ||    ||:| |||||  |:
Db    748  DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVTT   806

Qy    127  TNFVLDGTASADIPAGTYDYVIINPNP--GIYIVGEG---VSKGNDYVVEAGKTYHFTV   181
              :| ||  | :IPGG YDY I NP P        |  :|    |   | |||| :| |:
Db    807  QNIIVTGQGEVIPGGVYDYCITNPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFM    866

Qy    182  QRQGPGDAASVV-------                                          193
            :  | ||                                                    :
Db    867  RRAGMGDGTDMEVEDDSPASYTYVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT  926

Qy    194  --------VTGEGGNEFAPVQNLQWSVSGQTVLTWQAP-------ASDKRTYVLN      234
                    ||| ||||||||||| |:| |||:|||||:|       ||||||:||
Db    927  AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPGTTLS     986

Qy    235  ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD  294
            |||: | :|||| :|||||||| |  :  : |  ||: ||| :||: | :|| |||:|||
Db    987  ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042

Qy    295  NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS--  350
            ||| || : |||| || :||  | | | :|||  : |||||:| |||  ||||| ||
Db   1043  NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102

Qy    351  -------DKPAPMNLV--------                                     366
                   |||||:|:
Db   1103  RSPXAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy    367  --------PAPYQERTID--------LSAYAGQQV-                         385
                    |||:|||:||      |:|:
Db   1163  FESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY  1222

Qy    386  -------------------------------                             385
Db   1223  LISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFG 1282

Qy    386  --------------YLAFRHFNSTGIFRLYLDDV--AVSGEGGSNDYT            417
                          ||||||:||||||||||:||||   ||| | ||:|||
Db   1283  LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT 1342

Qy    418  YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV  477
            |||||| : ||  | ||||| :|| ||||:||||||||||||||: |||||| ||:||:|
Db   1343  YTVYRDGTNIKEGLTETTFEYGVATGNHEYCVEVKYTAGVSPKKCCVDVTV-NSTQFNPV 1401

Qy    478  QNLTGSAV-GQKVTLKWDAP                                         496
            |||| | : | :||:||:||
Db   1402  QNLTAEQAPNSMDAILKWNAP                                       1422
```

RESULT 9

```
O52050
ID   O52050        PRELIMINARY;   PRT;   1732 AA.
AC   O52050;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine specific cysteine protease.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=98298016; PubMed=9612563;
RA   Lewis J.P., Macrina F.L.;
RT   "IS195, an insertion sequence-like element associated with protease
RT   genes in Porphyromonas gingivalis.";
RL   Infect. Immun. 66:3035-3042(1998).
DR   EMBL; AF017059; AAC26523.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
```

RESULT 10

```
O07442
ID   O07442        PRELIMINARY;   PRT;   1732 AA.
AC   O07442;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine-specific cysteine proteinase.
GN   PRTK.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=99235907; PubMed=10219167;
RA   Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtK that encodes
RT   a lysine-specific cysteine proteinase and three sequence-related
RT   adhesins.";
RL   Oral Microbiol. Immunol. 14:92-97(1999).
DR   EMBL; U75366; AAB60809.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1732 AA;  187914 MW;  45D5B9137739I703 CRC64;

Query Match             30.5%;  Score 806.5;  DB 2;  Length 1732;
Best Local Similarity   29.7%;  Pred. No. 2.8e-41;
Matches  220;  Conservative  72;  Mismatches 172;  Indels 277;  Gaps  22;

Qy      22 TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD------------PIPAGMARIILEA 67
              ||:|||||                                      ||||||||||||
Db     693 TATTQGQKVTLKWEAPS-----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy      68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
              : :||| ||| ||:|:||| :|::|   |||||  ||||||| |||  ||||||||||
Db     748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 806

Qy     127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVBAGKTYHFTV 181
             | :::  : |       :||| || : :|: |   |     : :  ||:
Db     807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 866

Qy     182 QRQGPGDAASVV------------------------------------------------ 193
             ||||||||||
Db     867 RRAGMGDGTDMEVEDDSPASYTYTVVRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy     194 -------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP-----------ASDKRTYVLN 234
                         |  | :|| |||    :  : :|: |||
Db     927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS 986

Qy     235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
             |||:|||||||| |  |:|::  ::|| : :|  ||||||||||||||||||| ||||
Db     987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042

Qy     295 NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS--- 350
             :||: |   |||  | ||   :  ||:: :|       : |  ::|  :|    ||
Db    1043 NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102

Qy     351 -------------DKPAPMNLV-------------------------KSEGVKL------ 366
                         ||||||||                       :|:||:|
Db    1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy     367 ---------PAPYQERTID----------------------------LSAYAGQQV---- 385
                      ||||:||||                            |:|:|||
Db    1163 FESSTHGEAPAEWTTIDADGDGQGWLCLSSQOLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy     386 ------------------------------------------------------------ 385
                                                    ||:
Db    1223 LISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFRETPNGINKGGARFG 1282

Qy     386 ------------------YLAFRHFNSTGIFRIYLYLDDV--AVSGEGSSNDYT 417
                              || |:|:|  ||  ||  :|:||    ||  ||:|:|
Db    1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYT 1342
```

```
Qy     418 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
              ||||||||:||||| |  :|:|:||| |::| |||||||||||||||||||| :|| :
Db    1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFNPV 1401

Qy     478 QNLTGSAV--GQKVTLKWDAP 496
              ||||                :
Db    1402 QNLTAEQAPNSMDAILKWNAP 1422


RESULT 11
Q9ZNB5
ID   Q9ZNB5            PRELIMINARY;        PRT;  1223 AA.
AC   Q9ZNB5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   130k-HMGD (Fragment).
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=99143160; PubMed=9998746;
RA   Shibata Y., Hayakawa M., Takiguchi H., Shiroza T., Abiko Y.;
RT   "Determination and characterization of the hemagglutinin-associated
RT   short motifs found in Porphyromonas gingivalis multiple gene
RT   products.";
RL   J. Biol. Chem. 274:5012-5020(1999).
DR   EMBL; AB019363; BAA34341.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; PepPidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   NON_TER      1      1
SQ   SEQUENCE   1223 AA;  131542 MW;  00225CD2BA9F91B3 CRC64;

Query Match             30.5%;  Score 805.5;  DB 2;  Length 1223;
Best Local Similarity   29.6%;  Pred. No. 2e-41;
Matches  219;  Conservative  73;  Mismatches 172;  Indels 277;  Gaps  22;

Qy      22 TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD------------PIPAGMARILLEA 67
              ||:|||||                                      ||||||||| ||
Db     184 TATTQGQKVTLKWEAPS-----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 238

Qy      68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
              : :||| ||| ||:|:||| :|::|   |||||  ||||||| |||  ||||||||||
Db     239 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 297

Qy     127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVBAGKTYHFTV 181
             | :::  : |       :||| || : :|: |   |     : :  ||:
Db     298 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 357

Qy     182 QRQGPGDAASVV------------------------------------------------ 193
             ||||||||||
Db     358 RRAGMGDGTDMEVEDDSPASYTYTVVRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 417

Qy     194 -------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP-------ASDKRTYVLN 234
                         |  | :|| |||                      :|:
Db     418 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS 477

Qy     235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
             |||:|||||||| |  |:|::  ::|| : :|  ||||||||||||||||||| ||||
Db     235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
```

```
Query Match          21.3%;    Score 561.5;  DB 2;    Length 1097;
Best Local Similarity    29.8%;    Pred. No. 2.4e-26;
Matches 158;  Conservative  48;  Mismatches 117;  Indels 207;  Gaps  17;

Qy   165  KGNDYVVEAGKT--YHFT-------VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSG 214
               ||:||   :   |:||                      : :       ||| |||||||
Db   267  KGSDYTVATGRLRFGIDFTPEWSLNLYQNVFLGDAIPV----GGSNEFAPVQNLTGSSVG 322

Qy   215  QTVTLTWQAP--------ASDKRTYVLNESFDTQTLPNGWTMIDADGDHNWLSTINVY 265
               :||| ::|                :|||   | ||  |  |||:||||||||   :
Db   323  QKVTLKWDAPNGTPNPNPNPNPGTTLSESFE-NGIPASWKTIDADGDHGW-KPGNAP 380

Qy   266  NTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQ-VPWTNEH 324
               : :  ::: ::   |||| :|||:|:|||||  ||||||||||||||||||| :|:|||
Db   381  GIAGYNSNGCVYSRSF-GLGGIGV-LTPDNYLITPALDLPNGGKLTFWVCAQDANYASEH 438

Qy   325  YGVFLSTTGNEAANFTIKLLEETLGS----------------DKPAPMNLV------- 359
               :|:| |   ::  | ||||:|||||                  :|:||||||
Db   439  YAVTASSTGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFR 498

Qy   360  ------------------KSEGVKL------------------------PAPYQERTID- 376
                                 ||||||                          :|||||||||
Db   499  HFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSG 558

Qy   377  ----LSAYAGQQV----------------------------------------------- 385
                   :|:|:|
Db   559  QLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMI 618

Qy   386  ----------------------------------------------------YLAFR 390
                                                                    ||:||
Db   619  SKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFR 678

Qy   391  HFNSTGIFRLYLDDV--AVSGEGSSNDYTYTVRDNVVIAQNLAATTFNQENVAPGQYNY 448
               ||| :||||||:|||  ||| | ::|| ||| ::|| |  :  | :||| ||:|||||
Db   679  HYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEY 738

Qy   449  CVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAV--GQKVTLKWDAP 496
               |||||||||||||||||||||||| :|||||||||||     |||||||||
Db   739  CVEVKYTAGVSPKKCVVNTV-NSTQFNPVQNLTAEQAPNSMDAILKWNAP 787
```

RESULT 13

```
Q9KIB3
ID   Q9KIB3;              PRELIMINARY;      PRT;   312 AA.
AC   Q9KIB3;
DT   01-OCT-2000  (TrEMBLrel. 15, Created)
DT   01-OCT-2000  (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002  (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical outer membrane protein PG27.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C.;
RT   "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF237555; AAF81413.1; -.
SQ   SEQUENCE  312 AA;  34592 MW;  0D5792C9643A25F5 CRC64;

Query Match          14.4%;    Score 379;  DB 2;    Length 312;
Best Local Similarity    37.2%;    Pred. No. 8.6e-16;
Matches  94;  Conservative  44;  Mismatches  85;  Indels 30;  Gaps   9;
```

Second column (left):

```
Db   478  ESFE-NGIPASWKTIDADGDHGW-KPGNAPGIAGYNSNGCVYSRSF-GLGGIGV-LTPD 533
               |||:   :||:::   ||||||||                     :|:||:  |:||
Qy   295  NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS--- 350
               ||::|  |:   |||:|:|:|:: :  :|:  :|:| |       ::  |: |||||
Db   534  NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 593

Qy   351  ------------------DKPAPMNLV------------------------KSEGVKL-- 366
                                 :|:||||||                        ||||||
Db   594  RSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 653

Qy   367  -------PAPYQERTID------------------------------LSAYAGQV---- 385
                      :|||||||||                               |:|:|:|
Db   654  FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNY 713

Qy   386  ---------------------------------------------------------- 385
Db   714  LISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFG 773

Qy   386  -----------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
                                      ||:||||| :||||||||:|||  ||| | ::||||
Db   774  LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT 833

Qy   418  YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
               |||:||   |  :  | :||| ||:|||||||||||||||||||||||: :|||||::
Db   834  YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNPV 892

Qy   478  QNLTGSAV--GQKVTLKWDAP 496
               ||||:   ||||||||||||
Db   893  QNLTAEQAPNSMDAILKWNAP 913
```

RESULT 12

```
P72196
ID   P72196;              PRELIMINARY;      PRT;  1097 AA.
AC   P72196;
DT   01-FEB-1997  (TrEMBLrel. 02, Created)
DT   01-FEB-1997  (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003  (TrEMBLrel. 25, Last annotation update)
DE   TonB-linked adhesin precursor.
GN   TLA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97386416; PubMed=9244265;
RA   Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA   Curtis M.A.;
RT   "The tla gene of Porphyromonas gingivalis W50: a homologue of the
RT   arginine-specific protease precursor (PrpRI) which shares sequence
RT   similarity to TonB-linked receptors.";
RL   J. Bacteriol. 179:4778-4788(1997).
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR   EMBL; Y07618; CAA68897.1; -.
DR   GO; GO:0019867; C:outer membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Membrane; Outer membrane; Receptor; Signal; TonB box.
FT   SIGNAL        1     53       POTENTIAL.
SQ   SEQUENCE  1097 AA;  118731 MW;  73BA337B421F8B9 CRC64;
```

```
Qy  233 LNESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLS 292
Db    4 LSESFES-GIPAIWKTIDADGDGYNWMHLTN------FTGQSGLCVSSASYIGGVGA-LT 55

Qy  293 PDNYLVTPKVTVPENG--KLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLG 349
Db   56 PDNYLIITPELKLPTDALVEIIYWCTQDLTAPSEHYAVVSSSTGNNAADFVNLLYEETLT 115

Qy  350 SDKPAPMNLV---KSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVA 406
Db  116 AKRIQSPELIRGNRTQGVW----YQRKVV----LPNDTKYVAFRHFNSTDNFWLNLDEVS 167

Qy  407 V-------SGEGGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGV 458
Db  168 ILYTPLPRRAPCPHPGGYTYSVFRDGQKIASGLSALAYIDTDVPYGTQDYCVQVNYLQGD 227

Qy  459 SPKVCKDVTVEGS 471
Db  228 SYKVCKNIVVANS 240
```

RESULT 14
```
ID   Q9F4J0;    PRELIMINARY;    PRT;   925 AA.
AC   Q9F4J0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative outer membrane protein PG57.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA   Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT   "Identification of vaccine candidates from genomic analysis of
RT   Porphyromonas gingivalis.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY007522; AAG24228.1; -.
DR   InterPro; IPR003961; FN3; 1.
DR   SMART; SM00060; FN3; 1.
SQ   SEQUENCE   925 AA;  103632 MW;  5FF2198D6914DAE8 CRC64;

Query Match             12.8%;  DB 2;  Length 925;
Best Local Similarity   25.4%;  Score 337;  Pred. No. 1.6e-12;
Matches 152;  Conservative 62;  Mismatches 191;  Indels 194;  Gaps 31;

Qy    4 LNSLFSLAVLLSLLCW------------GQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKD 51
Db    5 MKSIVFRAFLTILLSWAAITNPTAQEISGMNASCLAAPAQPDTILYESFENG-------- 56

Qy   52 LRDPIPAGMARIILEAHDVWEDGTGYQNLWDADHNQYGASIPEESFWFANGT----IPA 106
Db   57 ---PVPNGWLEIDADA-----DGA----TWGSPSGSF--SVP----YGHNGLCTYSHIRS 98

Qy  107 GLYDPFEYKVPVNADA------------SFSPTNFVL---DGTASADIPAGTYDYVIIN 150
Db   99 GISTAGNYLITPNIEGAKRVKYWVCNQYSTNPEHYAVMVSTTGTAIEDFVLLFDDSITGK 158

Qy  151 PNPGI-----IYIVGEGVS-----KGNDYVVERAGKTYHFTVQR--QGPGDAASVVVTGE 197
Db  159 PTPLVWRRRIVDLPEGTKYIAWRHYKVTDSHTEFLKLDDVTVYRSIEGEPATDFTVINI 218
```

```
Qy  198 GGNEFAPVQNLQWS-------------------VSGQTVTLTWQAPASDK--------RTY 231
Db  219 GQN---VGRLTWNYPEDYQPEGKGNEEIQLSGYNIYANGTLLAQIKDVSILEYVDSTYS 274

Qy  232 -------------------------------------VLNESFDTQTLPNGWTMIDADGDGH 256
Db  275 LRDNPLQVEYCVTAVYDESIESSTVCGTLHYATDAILYENFENGPVPNGWLVIDADGDGF 334

Qy  257 NWLSTINVYNT-ATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGK-LSYWV 314
Db  335 SWGHYLNAYDAFPGHNGGHCSLSASYVPGIG--PVTPDNYLIITPKV--EGAKRVKYWV 388

Qy  315 SSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVKSEGVKLPAPYQBR 373
Db  389 STQDANWAAEHYAVMASTTGTAVGDFVI-LFEETM-TAKPT-------------GAWYER 433

Qy  374 TIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGGSSNDYTYTVYRDNVV--IAQNL 431
Db  434 TINLP--EGTK-YIAWRHYHCTDIYFLKLDDITVFGTPASEPEPVT---DFVVSLIENNK 487

Qy  432 AATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKD--VTVEGSNEFA----HVQNLT 481
Db  488 GRLKWN--------------------YPNGYEPDKTDKDPLQLAGYNIYANGSLLVHIQDPT 530
```

RESULT 15
```
ID   Q9XBU9;    PRELIMINARY;    PRT;    293 AA.
AC   Q9XBU9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Immunoreactive 32 kDa antigen PG25.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF155351; AAD38410.1; -.
SQ   SEQUENCE   293 AA;  32272 MW;  CC03EAC241F7F6F1 CRC64;

Query Match              8.1%;  Score 213.5;  DB 2;  Length 293;
Best Local Similarity   38.8%;  Pred. No. 1.5e-05;
Matches 52;  Conservative 19;  Mismatches 42;  Indels 21;  Gaps 4;

Qy  362 EGVKLPAPYQER-TIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGGSSNDYTYTV 420
Db   39 ESLTFVAPVEETDAIEAEVEALQEI----------------VATEEIA--EQAVRSYTYTV 81

Qy  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVBGSNEFAHVQNL 480
Db   82 YRDGVKIASGLTEPTFLDEDVPAGEHTYCVEVQYQGGVSDKVCVDVEVK---DFKPVTNL 138

Qy  481 TGSAVGQKVTLKWD 494
Db  139 TGTASNDEVSLDWD 152
```

Search completed: May 18, 2004, 11:35:05
Job time : 26.8443 secs

```
                    GenCore version 5.1.6
               Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        May 18, 2004, 11:30:39 ; Search time 8.24989 Seconds
                                        (without alignments)
                                        3110.116 Million cell updates/sec

Title:         US-08-353-485-2
Perfect score: 2641
Sequence:      1 MRKLNSLFSLAVLLSLLCWG..........QNLTGSAVGQKVTLKWDAPN 497

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:       389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     Issued Patents AA:*
               1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
               2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
               3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
               4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
               5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
               6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                               SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2641	100.0	497	2	US-08-570-311-2	Sequence 2, Appli
2	2641	100.0	497	2	US-08-353-485-2	Sequence 2, Appli
3	2641	100.0	2628	2	US-08-570-311-14	Sequence 14, Appli
4	951.5	36.0	1706	4	US-09-066-330-10	Sequence 10, Appli
5	944.5	35.8	1477	4	US-09-482-500A-1	Sequence 1, Appli
6	944.5	35.8	1704	3	US-08-336-308A-10	Sequence 10, Appli
7	944.5	35.8	1704	3	US-08-822-324-6	Sequence 6, Appli
8	944.5	35.8	1704	3	US-09-490-931-10	Sequence 10, Appli
9	943.5	35.7	1687	2	US-08-570-311-29	Sequence 29, Appli
10	816	30.9	1358	2	US-08-570-311-27	Sequence 27, Appli
11	812.5	30.8	1732	2	US-08-570-311-10	Sequence 10, Appli
12	812.5	30.8	1732	2	US-08-353-485-10	Sequence 10, Appli
13	806.5	30.5	1087	2	US-08-570-311-8	Sequence 8, Appli
14	806.5	30.5	1087	2	US-08-353-485-8	Sequence 8, Appli
15	806.5	30.5	1732	4	US-09-066-330-11	Sequence 11, Appli
16	546.5	20.7	450	2	US-08-570-311-16	Sequence 16, Appli
17	546.5	20.7	456	2	US-08-570-311-18	Sequence 18, Appli
18	546.5	20.7	456	2	US-08-570-311-20	Sequence 20, Appli
19	458.5	17.4	439	2	US-08-570-311-22	Sequence 22, Appli
20	405	15.3	942	1	US-08-141-324-14	Sequence 14, Appli
21	405	15.3	942	1	US-08-541-902-14	Sequence 14, Appli
22	175	6.6	49	3	US-08-822-324-18	Sequence 18, Appli
23	148	5.6	42	4	US-09-066-330-3	Sequence 3, Appli
24	140	5.3	2736	4	US-09-252-991A-30227	Sequence 30227, A
25	135.5	5.1	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
26	132	5.0	509	3	US-08-822-324-8	Sequence 8, Appli
27	127	4.8	46	3	US-08-822-324-9	Sequence 9, Appli
28	127	4.8	811	1	US-08-480-604A-7	Sequence 7, Appli
29	127	4.8	811	2	US-08-405-496A-7	Sequence 7, Appli
30	127	4.8	811	3	US-08-915-136-7	Sequence 7, Appli
31	127	4.8	811	4	US-08-957-310-7	Sequence 7, Appli
32	127	4.8	811	4	US-10-011-366-7	Sequence 7, Appli
33	127	4.8	811	4	US-09-084-517-7	Sequence 7, Appli
34	127	4.8	812	1	US-08-480-604A-29	Sequence 29, Appli
35	127	4.8	812	3	US-08-915-136-29	Sequence 29, Appli
36	127	4.8	812	4	US-09-084-517-29	Sequence 29, Appli
37	127	4.8	2710	1	US-08-480-604A-6	Sequence 6, Appli
38	127	4.8	2710	3	US-08-405-496A-6	Sequence 6, Appli
39	127	4.8	2710	3	US-08-915-136-6	Sequence 6, Appli
40	127	4.8	2710	4	US-08-957-310-6	Sequence 6, Appli
41	127	4.8	2710	4	US-10-011-366-6	Sequence 6, Appli
42	127	4.8	2710	4	US-09-084-517-6	Sequence 6, Appli
43	126	4.8	1752	4	US-09-865-621A-2	Sequence 2, Appli
44	123	4.7	951	4	US-09-924-097A-15	Sequence 15, Appli
45	121	4.6	25	2	US-08-902-516-46	Sequence 46, Appli

```
                               ALIGNMENTS

RESULT 1
US-08-570-311-2
; Sequence 2, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT:  Progulske-Fox, Ann
;   APPLICANT:  Tumwasorn, Somying
;   APPLICANT:  Lepine, Guylaine
;   APPLICANT:  Han, Naiming
;   APPLICANT:  Lantz, Marilyn
;   APPLICANT:  Patti, Joseph
;   TITLE OF INVENTION:  Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION:  and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES:  29
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Ted W. Whitlock
;     STREET:  2421 N.W. 41st Street, Suite A-1
;     CITY:  Gainesville
;     STATE:  FL
;     COUNTRY:  USA
;     ZIP:  32606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/570,311
;     FILING DATE:
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/353,485
;     FILING DATE:  09-DEC-1994
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/647,119
;     FILING DATE:  25-JAN-1991
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/241,640
;     FILING DATE:  08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Whitlock, Ted W.
;     REGISTRATION NUMBER:  36,965
;     REFERENCE/DOCKET NUMBER:  UF15.C3
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (904) 375-8100
;     TELEFAX:  (904) 372-5800
; INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
```

```
;        LENGTH: 497 amino acids
;        TYPE: amino acid
;        TOPOLOGY: linear
;        MOLECULE TYPE: protein
US-08-570-311-2

Query Match             100.0%;     Score 2641;  DB 2;  Length 497;
Best Local Similarity   100.0%;     Pred. No. 1.4e-205;  Indels 0;  Gaps 0;
Matches 497;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60

Qy   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120

Qy  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180

Qy  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240

Qy  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300

Qy  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDKPAPMNLVK 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDKPAPMNLVK 360

Qy  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420

Qy  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480

Qy  481 TGSAVGQKVTLKWDAPN 497
         |||||||||||||||||
Db  481 TGSAVGQKVTLKWDAPN 497

RESULT 2
US-08-353-485-2
; Sequence 2, Application US/08353485
; Patent No. 5830710
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION:  Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION:  and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES:  12
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Ted W. Whitlock
;     STREET:  2421 N.W. 41st Street, Suite A-1
;     CITY:  Gainesville
;     STATE:  FL
;     COUNTRY:  USA
;     ZIP:  32606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/353,485
;     FILING DATE:  09-DEC-1994
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/647,119
;     FILING DATE:  25-JAN-1991
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/241,640
;     FILING DATE:  08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Whitlock, Ted W.
;     REGISTRATION NUMBER:  36,965
;     REFERENCE/DOCKET NUMBER:  UF15.C2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (904) 375-8100
;     TELEFAX:  (904) 372-5800
INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 497 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-353-485-2

Query Match             100.0%;     Score 2641;  DB 2;  Length 497;
Best Local Similarity   100.0%;     Pred. No. 1.4e-205;  Indels 0;  Gaps 0;
Matches 497;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60

Qy   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120

Qy  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180

Qy  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240

Qy  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300

Qy  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDKPAPMNLVK 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDKPAPMNLVK 360

Qy  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420

Qy  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480

Qy  481 TGSAVGQKVTLKWDAPN 497
         |||||||||||||||||
Db  481 TGSAVGQKVTLKWDAPN 497

RESULT 3
US-08-570-311-14
; Sequence 14, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
```

```
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Ted W. Whitlock
;  STREET: 2421 N.W. 41st Street, Suite A-1
;  CITY: Gainesville
;  STATE: FL
;  COUNTRY: USA
;  ZIP: 32606
; COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/570,311
;  FILING DATE:
;  CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: US 08/353,485
;  FILING DATE: 09-DEC-1994
;  CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: US 07/647,119
;  FILING DATE: 25-JAN-1991
;  CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: US 07/241,640
;  FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
;  NAME: Whitlock, Ted W.
;  REGISTRATION NUMBER: 36,965
;  REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (904) 375-8100
;  TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 14:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 2628 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;  MOLECULE TYPE: protein
US-08-570-311-14
```

```
Query Match        100.0%;  Score 2641;  DB 2;  Length 2628;
Best Local Similarity  100.0%;  Pred. No. 1.7e-204;
Matches 497;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60

Qy     61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120

Qy    121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180

Qy    181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240

Qy    241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300
```

```
Qy    301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360

Qy    361 SEGVKLPAPYQERTIDLSAYAGQQVYLAPRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 SEGVKLPAPYQERTIDLSAYAGQQVYLAPRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420

Qy    421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480

Qy    481 TGSAVGQKVTLKWDAPN 497
            |||||||||||||||||
Db    481 TGSAVGQKVTLKWDAPN 497
```

```
RESULT 4
US-09-066-330-10
; Sequence 10, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
;  APPLICANT: Reynolds, Eric C.
;  APPLICANT: Bhogal, Peter S.
;  APPLICANT: Slakeski, Nada
;  TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;  FILE REFERENCE: Reynolds
;  CURRENT APPLICATION NUMBER: US/09/066,330A
;  CURRENT FILING DATE: 1998-09-15
;  EARLIER APPLICATION NUMBER: PN 6275
;  EARLIER FILING DATE: 1995-10-30
;  EARLIER APPLICATION NUMBER: PCT/AU96/00673
;  EARLIER FILING DATE: 1996-10-30
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;  LENGTH: 1706
;  TYPE: PRT
;  ORGANISM: Porphyromonas gingivalis
US-09-066-330-10

Query Match        36.0%;  Score 951.5;  DB 4;  Length 1706;
Best Local Similarity  31.9%;  Pred. No. 5.6e-68;
Matches 237;  Conservative  74;  Mismatches 156;  Indels 275;  Gaps  20;

Qy     22 TAAAQG--------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
            :||||          :.| ||| :|||||  ||    || ||.| ||| ||.|||||||:
Db    671 TATTQGQKVTLKWDAPSIKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
            ||| ||:|| |::|  ||.||  ||.|   :||||  | |||||||||| |||:||| |
Db    731 DVWNDGSGYQILLDADHDQYGQVIPSDIHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
            || :|.||:| ||||||  |::.||.|.| |  || |||.|| :::: ||.|| :|:||:
Db    790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy    186 PGDAASVV----------------------------------------------- 193
            ||:|.|||
Db    850 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909

Qy    194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS-----------DKRTYVLNE 235
                  |||.||| | |||||||::|| :| |||| ::::             :.|.||:|
Db    910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPDGTTTLSE 969

Qy    236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
            || ||| :  |:|.|::|::|||  |:|:: |::  |||   :|    || |:::| ||
Db    970 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1025

Qy    296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETL---GSD 351
            ||::| |  ||.||.|:|||:| |  .  |:|:|  ::| :|::  || ::| ::     :
Db   1026 YLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETTTAKGVR 1085
```

```
Qy   352  KPAPM------------------------------------------------------------  356
Db  1086  SPEAMRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF  1145
Qy   357  ----------------------------------------NLVKS------------------  361
                                                     :|:|
Db  1146  ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALNPDNYL  1205
Qy   362  ----------------------------------------------------------------  361

Db  1206  ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL  1265
Qy   362  ----EGVKLPAPYQERTIDLSAYAGQQVLAFRHFNSTGIFRLYLDDV--AVSGEGSSND  415
                 : :|| |||       || |   |||||    :|: :  ||  :
Db  1266  STEADGAKPQSWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD  1322
Qy   416  YTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA  475
               |||||| ||| ||||||     ||||  ||||||
Db  1323  YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN  1381
Qy   476  HVQNLTGSAVGQKVTLKWDAPN  497
               |::|:|
Db  1382  PVNLKAQPDGGDVVLKWEAPS  1403
```

```
Query Match         35.8%;   Score 944.5;   DB 4;   Length 1477;
Best Local Similarity   31.9%;   Pred. No. 1.7e-67;
Matches  235;  Conservative  75;  Mismatches 160;  Indels 267;  Gaps  20;

Qy    22  TRAAQG--------GPKTAPSVTHQAVQK--GIR----TSKVKDLRDPIPAGMARIILEAH   68
              ::|      ||        ||    |      :||:|:|||
Db   444  TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH  503
Qy    69  DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP  126
              ||:|||:|    ||||||       |::  |    ||||||
Db   504  DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW--PNCSVPANLFAPFEYTVPENADPSCCSP  562
Qy   127  TNFVLDGTASADIPAGTYDVIINPNPGI--IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG  185
              :|  |  :    |      :|::   :  :|::|
Db   563  TNM.IMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG  622
Qy   186  PGDAASVV-------------------------------------------------  193
              ||  |:|
Db   623  SGDGTELIISEGGGSDYTYTVYRDGTKIKEGLIATTFEEDGVATGNHEYCVEVKYTAGVS  682
Qy   194  ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTVLNESF  237
                    :|:|||||
Db   683  PKVCKDVIVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTLSESF  742
Qy   238  DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASSGAKIDLSPDNYL  297
```

Second column:

```
Db   743  E-NGIPASWKTIDADGDGHGW-KPGNAPGLAGYNSNGCVYSESF-GLGGIGV-LTPDNYL  798
               :  |||||||||||||                    :|:|:|    |||||
Qy   298  VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETLGS------  350
               :|| |:|:|: ||        ||||    ||  :||||
Db   799  ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP  858
Qy   351  -------------------------DKPAPMNLV--KSEGVKL-  366
                                        ||||| ||
Db   859  EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES  918
Qy   367  ---PAPYQERTID---------------------------LSAYAGQQV---------  385
               :|:|
Db   919  STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS  978
Qy   386  --------------------------------------------------------------  385

Db   979  KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST  1038
Qy   386  -----------------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV  420
                                         || |   |||||    :|: :
Db  1039  EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTV  1098
Qy   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
               |||| ||| ||||||     ||||  ||||||
Db  1099  YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL  1157
Qy   481  TGSAVGQKVTLKWDAPN  497
               |:|:|
Db  1158  KAQPDGGDVVLKWEAPS  1174
```

```
;  TITLE OF INVENTION:  IMMUNOGENIC COMPOSITIONS COMPRISING
;  TITLE OF INVENTION:  PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
;  TITLE OF INVENTION:  METHODS
;  NUMBER OF SEQUENCES:  24
;  CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;     STREET:  5370 Manhattan Circle, Suite 201
;     CITY:  Boulder
;     STATE:  CO
;     COUNTRY:  US
;     ZIP:  80303
;  COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/822,324
;     FILING DATE:  21-MAR-1997
;     CLASSIFICATION:  536
;  PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 60/013,945
;     FILING DATE:  22-MAR-1996
;  ATTORNEY/AGENT INFORMATION:
;     NAME:  Ferber, Donna M.
;     REGISTRATION NUMBER:  33,878
;     REFERENCE/DOCKET NUMBER:  103-95 WO
;  TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (303) 488-8080
;     TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  6:
;  SEQUENCE CHARACTERISTICS:
;     LENGTH:  1704 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-822-324-6

Query Match             35.8%;    Score 944.5;  DB 3;  Length 1704;
Best Local Similarity   31.9%;    Pred. No. 2.1e-67;
Matches  235;  Conservative   75;  Mismatches 160;  Indels 267;  Gaps   20;

Qy     22 TAAAQG-------GPKTAPSVTHQAVQK--GIR---TSKVKLRDPIPAGMARIILEAH 68
Db    671 TATTQGGKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
Db    731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
Db    790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy    186 PGDAASVV-------------------------------------------------- 193
Db    850 SGDGTELTISEGGGSDYTYTVRDGTKIKEGLTATFEEDGVATGNHEYCVEVKYTAGVS 909

Qy    194 ------VTGEGNEFAPVQNLQWSVSGQTVLTWQAPAS-------------DKRTYVLNESF 237
Db    910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSESF 969

Qy    238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
Db    970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy    298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS------ 350
Db   1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy    351 -----------DKPAPMNLV------------------------------KSEGVKL----- 366
Db   1086 EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145
```

```
;  INFORMATION FOR SEQ ID NO:  10:
;  SEQUENCE CHARACTERISTICS:
;     LENGTH:  1704 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-336-308A-10

Query Match             35.8%;    Score 944.5;  DB 3;  Length 1704;
Best Local Similarity   31.9%;    Pred. No. 2.1e-67;
Matches  235;  Conservative   75;  Mismatches 160;  Indels 267;  Gaps   20;

Qy     22 TAAAQG-------GPKTAPSVTHQAVQK--GIR---TSKVKLRDPIPAGMARIILEAH 68
Db    671 TATTQGGKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
Db    731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
Db    790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy    186 PGDAASVV-------------------------------------------------- 193
Db    850 SGDGTELTISEGGGSDYTYTVRDGTKIKEGLTATFEEDGVATGNHEYCVEVKYTAGVS 909

Qy    194 ------VTGEGNEFAPVQNLQWSVSGQTVLTWQAPAS-------------DKRTYVLNESF 237
Db    910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSESF 969

Qy    238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
Db    970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy    298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS------ 350
Db   1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy    351 -----------DKPAPMNLV------------------------------KSEGVKL----- 366
Db   1086 EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145

Qy    367 ----PAPYQERTID------------------------------LSAYAGQQV------- 385
Db   1146 STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy    386 -------------------------------------------------------- 385
Db   1206 KDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGLST 1265

Qy    386 -----------YLAFRHFNSTGIFRLYLDDV--AVSGEGGSNDYTYTV--------- 420
Db   1266 EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNVILLDDIQFTMGGSPTPTDYTYTV 1325

Qy    421 YRDNVVIAQNLAATTNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
Db   1326 YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NFTQFNPVKNL 1384

Qy    481 TGSAVGQKVTLKWDAPN 497
Db   1385 KAQPDGDGVVLKWEAPS 1401
```

```
RESULT 7
US-08-822-324-6
; Sequence 6, Application US/08822324
; Patent No. 6129917
;  GENERAL INFORMATION:
;     APPLICANT:  Potempa, Jan S.
;     APPLICANT:  Travis, James
;     APPLICANT:  Genco, Caroline A.
```

```
Matches 235;  Conservative 75;  Mismatches 160;  Indels 267;  Gaps 20;

Qy    22  TAAAQG------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH   68
Db   671  TATTGQKVTLKWDAPSTKTNATTNTARSVDGIRELVILLSVSDAPELLRSGQAEIVLEAH  730

Qy    69  DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP  126
Db   731  DVWNDGSGYQILLDADHDQYGQVIPSDTHTIW-PNCSVPANLFAPFEYTVPENADPSCSP  789

Qy   127  TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG  185
Db   790  TNMIMDGTASVNIPAGTYDFSAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG  849

Qy   186  PGDAASVV-------                                               193
Db   850  SGDGTELTISEGGSGDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS  909

Qy   194  -----VTGEGNEFAPVQNLQWSVSGQTVTLTWQAPAS------------DKRTYVLNESF  237
Db   910  PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESF    969

Qy   238  DTQTLPNGWTMIDADGDGHNWLSTNVYNTATHTGDGAMPSKSWTASGGAKIDLSPDNYL   297
Db   970  E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy   298  VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETLGS-------  350
Db  1026  ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy   351  ------------DKPAPMNLV------------KSBGVKL-                     366
Db  1086  EAIRGRIQGTWRQKTVDLPAGTKYYAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145

Qy   367  ---PAPYQERTID-------------LSAYAGQQV-------                    385
Db  1146  STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy   386  ------------                                                  385
Db  1206  KDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST  1265

Qy   386  --------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV           420
Db  1266  EANGAKPQSVWIERTVDLPAGTKYYAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTV 1325

Qy   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
Db  1326  YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1384

Qy   481  TGSAVGQKVTLKWDAPN                                             497
Db  1385  KAQPDGGDVVLKWEAPS                                            1401
```

RESULT 8
US-09-490-931-10
; Sequence 10, Application US/09490931
; Patent No. 6274718
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/490,931
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/336,308
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-490-931-10

Query Match   35.8%;   Score 944.5; DB 3; Length 1704;
Best Local Similarity 31.9%;   Pred. No. 2.1e-67;

RESULT 9
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville

```
Qy    367 ------PAPYQERTID------LSAYAGQQV--                              385
          :                    :.:
Db   1127 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNVL 1186

Qy    386 ------                                                         385
Db   1187 ISKDVTGATKVKYYAVNDGFPGDHYAMISKTGTNAGDFTVFEETPNGINKGGARFGL    1246

Qy    386 ------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTY                      418
                :|||||:|:  |||||  :|||:||          ||||
Db   1247 STEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTY  1306

Qy    419 TYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ   478
          |:||. ::|.|| :|::||||:||.:|||:|||||||||||:|.   |:|:|:||| |:
Db   1307 TVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVK   1365

Qy    479 NLTGSAVGQKVTLKWDAPN                                           497
          ||   :  |.|||||:||.
Db   1366 NLKAQPDGGDVVLKWEAPS                                          1384
```

RESULT 10
US-08-570-311-27
; Sequence 27, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Arn
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 27:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1358 amino acids

;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 29:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1687 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-29

Query Match   35.7%; Score 943.5; DB 2; Length 1687;
Best Local Similarity   31.8%; Pred. No. 2.5e-67;
Matches 235; Conservative 75; Mismatches 160; Indels 269; Gaps 20;

```
Qy     22 TAAAQG---------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH   68
          || :||          |::.|.:.   : :  |||    :|.||| : :||||||:|:|||
Db    652 TATTQGGKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH   711

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP  126
          |||:||:||| ||||||:||| ::|    .  .||  .|||:| || :.|.||||: ||
Db    712 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFBYTVPENADPSCSP  770

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVCRQG  185
          || ::|||||::|||||||:.|  |..  .:|| || : : ||||:|||| |||:: ::|
Db    771 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG  830

Qy    186 PGDAASVV---------                                             193
          .||:
Db    831 SGDGTELTISEGGSDYTYTVRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS   890

Qy    194 --------VTGEGNEFAPVQNLQWSVSGQTVTLTWQAPAS----------DKRTYVLNE  235
                  ||:|||||||||||: .||.||.|||.|||:           |.||  ::
Db    891 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSE    950

Qy    236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN  295
          || .: :|.| |.|||||||| ||  :    :.   |.|. : :   ..|| .|||:|||
Db    951 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1006

Qy    296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS----  350
          || |.  ::|:|||| :|| :  | . ||:|.| ::||| :|:||| .|||||: :
Db   1007 YLITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR  1066

Qy    351 ------------DKPAPMNLV------KSEGVKL------                      366
                            :::      :|||.::
Db   1067 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1126
```

;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-27

Query Match   30.9%; Score 816; DB 2; Length 1358;
Best Local Similarity   29.6%; Pred. No. 3.7e-57;
Matches 219; Conservative 75; Mismatches 171; Indels 274; Gaps 21;

```
Qy    22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARILEA      67
            ||||         :|:   ::  :  :|:|          :|:||  |  |:|:|
Db   328 TATTQGQKVTLKWDAPS-----AKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAA 382

Qy    68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
         |:||  ||||| | |||||| ||::|||: || |   | |:|: :||:| || |:  ::
Db   383 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLLPANADPVVTT 441

Qy   127 TNFVLDGTASADIPAGTYDVIINPNP--GIIYIVGEG----VSKGNDYVVEAGKTYHFTV 181
          | :: | : |:|| |:||  |:|:|  ||  ||| |    |::|||: :|||||: |||
Db   442 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGDNQPARYDDFTFEAGKKYTFTM 501

Qy   182 QRQGPGDAASVV--                                              193
          | ||| :|:||
Db   502 RRAGMGDGTDMEVEDDSPASYTYIVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 561

Qy   194 ------VTGEGNEFAPVQNLQWSVSGQTVTLTWQAPAS-----DKRTYVLNE         235
               ||| |||||||||:  : :|| ||:||| |   ::      |:|: ||:
Db   562 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTLSE 621

Qy   236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
         ||:   :| | |  |||||||| | : : : |:|  |: | :  ||:  |  | ||:|| 
Db   622 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 677

Qy   296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
         ||: | : : |||||::||  |  ::  |||| : :|:||:|:||| | |||| |
Db   678 YLIIPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 737

Qy   351 ---------------DKPAPMNLV----------------------KSEGVKL------ 366
                        :|:   ||                       ||:|::
Db   738 SPEAIRGRIOGTWRQKTVDLIPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 797

Qy   367 ---PAPYQERTID--------------------LSAYAGQQV--                 385
            | |:|:| |                    |||:|| |
Db   798 ESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGTNVVASFSWNGMAINPDNYL 857

Qy   386 ---------------------------------------------               385
                
Db   858 ISKDVTGATKVKYYYAVNDGPPGDHYAVMLSKTGTNAGDFTVVFEETPNGINKGGARFGL 917

Qy   386 -------------YLAFRHFNSTGIFRIYLDDV--AVSGEGSSNDYTY             418
                      ||||||| | :| ||:|||:|   ||:|| ||||||
Db   918 STEANGAKPQSVWIERTVDLIPAGTKYVAFRHYNCSDLDYILLDDIQFTMGGSPTPTDYTY 977

Qy   419 TVYRDNVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ  478
         ||||| : :|:| :||| |::||:| |:| |||||||||||||||  :||| |  || |
Db   978 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNPVK 1036

Qy   479 NLITGSAVGQKVTLKWDAPN                                         497
         ||:   :||| |||||:||
Db  1037 NLKAQPDGGDVVLKWEAPS                                         1055
```

RESULT 11
US-08-570-311-10
; Sequence 10, Application US/08570311
; Patent No. 5824791
GENERAL INFORMATION:
  APPLICANT: Progulske-Fox, Ann
  APPLICANT: Tumwasorn, Somying
  APPLICANT: Lepine, Guylaine
  APPLICANT: Han, Naiming
  APPLICANT: Lantz, Marilyn
  APPLICANT: Patti, Joseph

;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1732 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-10

Query Match   30.8%; Score 812.5; DB 2; Length 1732;
Best Local Similarity   30.0%; Pred. No. 1e-56;
Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;

```
Qy    22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARILEA      67
            ||||         :|:   ::  :  :|:|          :|:||  |  |:|:|
Db   693 TATTQGQKVTLKWEAPS----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy    68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
         |:||  ||||| | |||||| ||::|||: || |   | |:|: :||:| || |:  ::
Db   748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLLPANADPVVTT 806

Qy   127 TNFVLDGTASADIPAGTYDVIINPNP--GIIYIVGEG----VSKGNDYVVEAGKTYHFTV 181
          | :: | : |:|| |:||  |:|:|  ||  ||| |    |::|||: :|||||: |||
Db   807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 866

Qy   182 QRQGPGDAASVV--                                              193
          | ||| :|:||
Db   867 RRAGMGDGTDMEVEDDSPASYTYIVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy   194 ------VTGEGNEFAPVQNLQWSVSGQTVTLTWQAP-----ASDKRTYVLN          234
               ||| |||||||||:  : :|| ||:||| |      ::  |:|: ||
Db   927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS  986

Qy   235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
         ||:   :| | |  |||||||| | : : : |:|  |: | :  ||:  |  | ||:|
Db   987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042
```

```
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 1732 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-353-485-10

Query Match             30.8%;  Score 812.5;  DB 2;  Length 1732;
Best Local Similarity   30.0%;  Pred. No. 1e-56;
  Matches  222;  Conservative   70;  Mismatches  172;  Indels  277;  Gaps   22;

Qy   22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD-------PIPAGMARILEA 67
            ||      :|:  :   ::||
Db  693 TATTCGQKVTLKWEAPS----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy   68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
Db  748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVTT 806

Qy  127 TNFVLDGTASADIPAGTYDVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKYHFTV 181
Db  807 QNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFM 866

Qy  182 QRQGPGDAASVV------ 193
Db  867 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy  194 -------VTGEGGNEFAPVQNLQWSVSGQTVLTWQAP-------ASDKRTYVLN 234
Db  927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS 986

Qy  235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTDGGAMFSKSWTASGGAKIDLSPD 294
Db  987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042

Qy  295 NYLIVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS-- 350
Db 1043 NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEETITAKGV 1102

Qy  351 ------DKPAPMNLV-------------KSEGVKL----- 366
Db 1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy  367 ------PAPYQERTID-------------LSAYAGQQV------ 385
Db 1163 FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy  386 -------- 385
Db 1223 LISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFG 1282

Qy  386 ------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
Db 1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT 1342

Qy  418 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
Db 1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTV-NSTQFNPV 1401

Qy  478 QNLTGSAV--GQKVTLKWDAP 496
Db 1402 QNLTAEQAPNSMDAILKWNAP 1422
```

```
Qy  295 NYLIVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS-- 350
Db 1043 NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEETITAKGV 1102

Qy  351 ------DKPAPMNLV-------------KSEGVKL----- 366
Db 1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy  367 ------PAPYQERTID-------------LSAYAGQQV------ 385
Db 1163 FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy  386 -------- 385
Db 1223 LISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFG 1282

Qy  386 ------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
Db 1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT 1342

Qy  418 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
Db 1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTV-NSTQFNPV 1401

Qy  478 QNLTGSAV--GQKVTLKWDAP 496
Db 1402 QNLTAEQAPNSMDAILKWNAP 1422
```

```
Db   371  GW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYLITPALDiANGGKITFWVCA 427
          |||||::|:::|::||||||||::||:::|     ||        ::||:|||:|| :::|    --DKP 353
Qy   317  Q-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS-
          ::  :|||:|:|:|||:|||||| ||||:|
Db   428  QDANYASEHYAVVASSTGNDASNFTNALLEETITAKGVRSPEAIRGRIOGTWROKTVDLP 487

Qy   354  APMNLV-----------------------KSBGVKL-----------PAPYQERTID----- 376
          ::|||:|             |:|         :|
Db   488  AGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDG 547

Qy   377  ----------------LSAYAGQQV-- 385
                          ||:|
Db   548  QDWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGF 607

Qy   386  ---------------------------------- 385
Db   608  PGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLP 667

Qy   386  ----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQE 439
          ::||:|        :|||||||||||||||:    |||
Db   668  AGTKYVAFRHYNCSDLDYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEED 727

Qy   440  NVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQKVTLKWDAPN 497
          :|: ||||||||||||||||||||||:    |:|  :|    || :|:
Db   728  GVATGNHEYCVEVKYTAGVSPKVCVNVTi-NPTQFNPVKNLKAQPDGGDVLKWEAPS 784
```

RESULT 14

```
US-08-570-311-8

Query Match          30.5%;  Score 806.5;  DB 2;  Length 1087;
Best Local Similarity 29.4%;  Pred. No. 1.6e-56;
Matches 211;  Conservative 76;  Mismatches 166;  Indels 265;  Gaps 19;

Qy    39  AVQKGIRTSKVKDLRD----------PIPAGMARIILEAHDVWEDGTGYQMLWDADHNQY 88
          :::||:||| :|||          |::|  |    |:||||||:  :|     ||||
Db    73  SAKKAEASREVKRIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYFLLDADHNTF 132

Qy    89  GASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYV 147
          |:|:|| |:     || |||||    ||| :|   |  :  :  :|||
Db   133  GSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVYDYC 191

Qy   148  IINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTVQRQPGDAASVV--------- 193
          ::|:|     ::|:|:     |:|::|||||||||:|||
Db   192  ITNPEPASGKMWIAGDGDNQPARYDDFTPEAGKKYTFTMRRAGMGDGTDMEVEDDSPASY 251

Qy   194  -----------------------------------VTGEGGNEFAP 204
                                             |||:||||||
Db   252  TYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAP 311

Qy   205  VQNLQWSVSGQTVTLTWQAPAS-----------DKRTVLNESFDTQTLPNGWTMIDADGDGH 256
          ||||   |:|:||  ||         :|:    |:::
Db   312  VQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSESFE-NGIPASWKTIDADGDGH 370

Qy   257  NWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYWVSS 316
```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1087 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-353-485-8

```
Query Match         30.5%;   Score 806.5;  DB 2;   Length 1087;
Best Local Similarity   29.4%;   Pred. No. 1.6e-56;
Matches  211;  Conservative   76;  Mismatches 166;  Indels 265;  Gaps   19;

Qy    39 AVQKGIRTSKVKDLRD---------PIPAGMARIILEAHDVWEDGTGYQMLWDADHNQY  88
            :::||| | |:::|        |||   :||||| ||||||||
Db    73 SAKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTF 132

Qy    89 GASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYV 147
            ||:| |:: || :| : ||  |||||| :|::::||||||||||
Db   133 GSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVYDYC 191

Qy   148 IINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFV---------VTGEGGNEFAP 193
            ||:||  ::||:|||    :||   || |||||| :
Db   192 ITNPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASY 251

Qy   194 ------------------VTGEGGNEFAP 204
                             |||||||||
Db   252 TYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYSPKVCKDVTVEGSNEFAP 311

Qy   205 VQNLQWSVSGQTVTLTWQAPAS-----DKRTYVLNESFDTQTLPNGWTMIDADGDGH 256
            ||||      |:|||||||        :|:|||||  :::: :|||||||||||
Db   312 VQNLTGSAVGQKVTLKWDAPNGTPNPNPGTTLSESFE-NGIPASWKTIDADGDGH 370

Qy   257 NWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYWVSS 316
            :: |:||||:||||||||||||||||| |||||: :: : : :|||
Db   371 GW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYLITPALDLANGGKLITFWVCA 427

Qy   317 Q-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS-------------DKP 353
            : :||||||||||||||  |||   |||||::
Db   428 QDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVAVYASSTGNDASNFTNALLEE 487

Qy   354 APMNLV------------------KSEGVKI----------- 376
            :|:|::            : :||:
Db   488 AGTKYYAVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDG 547

Qy   377 ------------LSAYAGQQV----------- 385
                       |::|||::
Db   548 QDWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNVLISKDVTGATKVKYYYAVNDGF 607

Qy   386 --------------- 386
Db   608 PGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLP 667

Qy   386 ---------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 439
                          ::|: ||| :||||||| :::
Db   668 AGTKYYAVAFRHYNCSDLDYILLDDIQFMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEED 727

Qy   440 NVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQKVTLKWDAPN 497
            ||: |  ::|:: ||||||||||||||||||:
Db   728 GVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNFPVKNLKAQPDGGDVVLKWEAPS 784
```

RESULT 15
US-09-066-330-11
; Sequence 11, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275

; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
; LENGTH: 1732;
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-11

```
Query Match          30.5%;    Score 806.5;   DB 4;    Length 1732;
Best Local Similarity   29.7%;   Pred. No. 3.2e-56;
Matches  220;  Conservative  72;  Mismatches 172;  Indels 277;  Gaps   22;

Qy    22 TAAAQGGPKI---APSVTHQAVQKGIRTSKVKDLRD------------PIPAGMARIILEA  67
            ||:   ::   ::: | ::::::|||                |||||||||||
Db   693 TATTQGQKVTLKWEAPS---AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy    68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
            |||||||||| :||||||| ||:| |:: ||  :| : ||  |||||| :
Db   748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 806

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG--VSKGNDYVVEAGKTYHFV 181
            :||||||||||||||||||||||:||  ::||:|||   :||   || |||||| :
Db   807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFTM 866

Qy   182 QRQGPGDAASVV--------------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP 234
            :  :                        :|||||||||||||    |:|||||||
Db   867 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy   235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
            ||:|: :|||||||||||||  :: |:||||:||||||||||||||||| |||||:
Db   927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPGTTLS 986

Qy   295 NYILVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS-- 350
            ||| : :: :|||  : :||: : :||||||||||||||  |||   |||||::
Db   987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042

Qy   351 ---------KSEGVKL----------- 366
                    : :||:
Db  1043 NYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102

Qy   367 --------PAPYQERTID----------LSAYAGQQV------ 385
            :  :|||:|: :::|:| :::|||::
Db  1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYYAVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy   386 --------------- 385
Db  1163 FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy   386 --------------- 385
Db  1223 LISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFG 1282

Qy   386 ---------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
                          ::|: ||| :||||||| :::
Db  1283 LSTEANGAKPQSVWIERTVDLPAGTKYYAVAFRHYNCSDLNYILLDDIQFMGGSPTPTDYT 1342

Qy   418 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
            ||||:||:  ::|| ||| :|: ||:|:::|:: ||||||||||||||||||||:
Db  1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFNPV 1401

Qy   478 QNLTGSAV--GQKVTLKWDAP 496
            |||| ||  ||:|||||||
Db  1402 QNLTAEQAPNSMDAILKWNAP 1422
```

Search completed: May 18, 2004, 11:36:47
Job time : 13.2499 secs

OM protein - protein search, using sw model

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2004  Compugen Ltd.
```

Run on: May 18, 2004, 11:32:49 ; Search time 19.8443 Seconds
                 (without alignments)
                 6969.043 Million cell updates/sec

Title: US-08-353-485-2
Perfect score: 2641
Sequence: 1 MRKLNSLFSLAVLLSLLCWG........QNLTGSAVGQKVTLKWDAPN 497

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5
Searched: 1145568 seqs, 277261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications AA:*
  1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
  2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
  3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
  4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
  5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
  6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
  7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
  8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
  9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	951.5	36.0	1706	14	US-10-229-066-10	Sequence 10, Appl
2	806.5	30.5	1732	14	US-10-229-066-11	Sequence 11, Appl
3	791.5	30.0	419	15	US-10-174-695-3	Sequence 3, Appli
4	662.5	25.1	419	15	US-10-174-695-5	Sequence 5, Appli
5	457.5	17.3	231	15	US-10-174-695-6	Sequence 6, Appli
6	188.5	7.1	196	15	US-10-174-695-4	Sequence 4, Appli
7	150	5.7	29	15	US-10-387-977-27	Sequence 27, Appl
8	150	5.7	1483	12	US-10-282-122A-51483	Sequence 51483, A
9	148	5.6	42	14	US-10-229-066-3	Sequence 3, Appli
10	145	5.5	509	15	US-10-387-977-101	Sequence 101, App
11	144	5.5	31	15	US-10-387-977-6	Sequence 6, Appli
12	143	5.4	1946	12	US-10-282-122A-62947	Sequence 62947, A
13	141	5.3	31	15	US-10-387-977-9	Sequence 9, Appli
14	139	5.3	26	15	US-10-387-977-15	Sequence 15, Appl
15	138.5	5.2	2468	12	US-10-282-122A-66335	Sequence 66335, A
16	138.5	5.2	2468	14	US-10-246-330-4	Sequence 4, Appli
17	137	5.2	25	15	US-10-387-977-82	Sequence 82, Appl
18	137	5.1	579	15	US-10-369-493-9075	Sequence 9075, Ap
19	134	5.1	2435	12	US-10-282-122A-47453	Sequence 47453, A
20	132	5.0	698	14	US-10-156-761-11953	Sequence 11953, A
21	129	4.9	973	14	US-10-156-761-9394	Sequence 9394, Ap
22	128	4.8	26	15	US-10-387-977-14	Sequence 14, Appl
23	127	4.8	811	12	US-10-272-898-7	Sequence 7, Appli
24	127	4.8	811	14	US-10-011-366-7	Sequence 7, Appli
25	127	4.8	811	15	US-10-354-774-7	Sequence 7, Appli
26	127	4.8	811	15	US-10-271-012-7	Sequence 7, Appli
27	127	4.8	2710	12	US-10-272-898-6	Sequence 6, Appli
28	127	4.8	2710	14	US-10-011-366-6	Sequence 6, Appli
29	127	4.8	2710	15	US-10-354-774-6	Sequence 6, Appli
30	127	4.8	2710	15	US-10-271-012-6	Sequence 6, Appli
31	126	4.8	1752	14	US-10-387-388-2	Sequence 2, Appli
32	125.5	4.8	577	15	US-10-369-493-21739	Sequence 21739, A
33	124	4.7	2013	12	US-10-282-122A-60608	Sequence 60608, A
34	124	4.7	2358	12	US-10-282-122A-45763	Sequence 45763, A
35	123.5	4.7	1204	12	US-10-282-122A-49627	Sequence 49627, A
36	123	4.7	866	14	US-10-222-038-2	Sequence 2, Appli
37	123	4.7	951	9	US-09-924-097-15	Sequence 15, Appl
38	122	4.6	806	15	US-10-369-493-3678	Sequence 3678, Ap
39	121	4.6	25	9	US-09-847-185-46	Sequence 46, Appl
40	121	4.6	25	10	US-09-930-915A-18	Sequence 18, Appl
41	121	4.6	25	14	US-10-224-286-46	Sequence 46, Appl
42	121	4.6	25	14	US-10-082-014-40	Sequence 40, Appl
43	121	4.6	25	14	US-10-372-076-41	Sequence 41, Appl
44	121	4.6	25	15	US-10-387-977-76	Sequence 76, Appl
45	121	4.6	25	15	US-10-387-977-79	Sequence 79, Appl

ALIGNMENTS

RESULT 1
US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US2003015763_7A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match   36.0%; Score 951.5; DB 14; Length 1706;
Best Local Similarity   31.9%; Pred. No. 3.5e-75;
Matches 237; Conservative 74; Mismatches 156; Indels 275; Gaps 20;

Qy   22 TAAAQG------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
                            | | : | : | : |: :| |:| ::::||
Db   671 TATTGGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
              |||:|||:||||::||::| : | | : : ||||| :::|||||: |:|
Db   731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789
```

Matches 220; Conservative 72; Mismatches 172; Indels 277; Gaps 22;

```
Qy    22 TAAAQGPKT---APSVTHQAVQKGIRTSKVKDLRD------PIPAGMARIILEA   67
         :      :||   . :   .   | : | :  :   |:|         |:|:
Db   693 TATTQGQKVTLKWEAPS-----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA   747

Qy    68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP   126
         . :|.|:|| |:|:|||||:.:||.||      | ||  |.|    ||| :| |||.::|
Db   748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVTT    806

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV   181
         .|:: :|| .  ||||.||. | ||:|    : | |.    :.| ..  :||||:|:||:
Db   807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQOPARYDFTFBAGKKYTFTM   866

Qy   182 QRQGPGDAASVV--                                                193
         : :|  |||:||
Db   867 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT   926

Qy   194 --------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP----------ASDKRTYVLN    234
                 |.. |.||||||||||: .|| |||:||| ||||           ||.|||||:
Db   927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS    986

Qy   235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD   294
         |||   | ||| ||||||||||.| |:|  ||| :||: |  . |:   |..| |:|:||
Db   987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD  1042

Qy   295 NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---   350
         ||| || :.:|:|||| :|| .| |.:..|||.|: |:||:.|.|||  ||||||
Db  1043 NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETTTAKGV  1102

Qy   351 -------------------KSEGVKL----                                366
                            : :  : |
Db  1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET  1162

Qy   367 -------PAPYQERTID--------------LSAYAGQQV-----                 385
                . | :|||                  | |.|.||
Db  1163 FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY  1222

Qy   386 ---------------DKPAPMNLV-------                               385
                        :|. | .:|
Db  1223 LISKDVTGATKVKYYYAVNDGPFGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFG  1282

Qy   386 ----------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT                  417
                   ||||||:|| |||||||||||  ||||||||:|||
Db  1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT  1342

Qy   418 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV   477
         ||||||. .|:|||  |||::|||| |.:|:|||||||||||||.|.: || .|.:| .|
Db  1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQPNV  1401

Qy   478 QNLTGSAV--GQKVTLKWDAP                                         496
         ||||     .| ||| :|:||
Db  1402 QNLTAEQAPNSMDAIIKWNAP                                        1422
```

```
Qy   127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG   185
         .|:: :|| |.:||||||||:.|    |  : | |.  .|.||||:||||.|||::.  |
Db   790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG   849

Qy   186 PGDAASVV--                                                    193
         .||.|:|
Db   850 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS   909

Qy   194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS----DKRTYVLNE          235
               |.. |.|||||||||| .|| |||:|||.||||      ||:||||:|
Db   910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSE      969

Qy   236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN  295
         ||   | ||| ||||||||||.| |:|  ||| :||: |  . |:   |..| |:|:|||
Db   970 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN  1025

Qy   296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETL---GSD   351
         ||| || :.:|:|||| :|| .| |.:..|||.|: |:||:.|.|||  ||||||
Db  1026 YLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1085

Qy   352 KPAPM                                                         356
         :
Db  1086 SPEAMRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF  1145

Qy   357 ----------NLVKS------                                         361
                   .|:|:
Db  1146 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALNPDNYL  1205

Qy   362 -------------------                                           361
Db  1206 ISKDVTGATKVKYYYAVNDGPFGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL  1265

Qy   362 ---EGVKLPAPYQERTIDLSAYAGGQVYLAFRHFNSTGIFRLYLDDV--AVSGEGSSND   415
            |.||| .| :||||.|.|.||.||:|||||:|| |||||||||||  ||||||||:
Db  1266 STEADGAKPQSVWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD  1322

Qy   416 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA    475
         ||| |||. .|:|||  |||::|||| |.:|:|||||||||||||.|.: || .|.:|
Db  1323 YTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN  1381

Qy   476 HVQNLTGSAVGQKVTLKWDAPN                                        497
         .:||||   .| ||| :|:||:
Db  1382 PVKNLKAQPDGGDVVLKWEAPS                                       1403
```

RESULT 2
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20031015763741
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric C.
;   APPLICANT: Bhogal, Peter S.
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;   FILE REFERENCE: Reynolds
;   CURRENT APPLICATION NUMBER: US/10/229,066
;   CURRENT FILING DATE: 2002-08-28
;   PRIOR APPLICATION NUMBER: US/09/066,330
;   PRIOR FILING DATE: 1998-09-15
;   PRIOR APPLICATION NUMBER: PN 6275
;   PRIOR FILING DATE: 1995-10-30
;   PRIOR APPLICATION NUMBER: PCT/AU96/00673
;   PRIOR FILING DATE: 1996-10-30
;   NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 11
;   SOFTWARE: PatentIn Ver. 2.0
;   LENGTH: 1732
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

Query Match   30.5%;   Score 806.5; DB 14; Length 1732;
Best Local Similarity 29.7%; Pred. No. 3.2e-62;

RESULT 3
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: Slakeski, Nada
;   APPLICANT: Chen, Chao Guang
;   APPLICANT: Barr, Ian George
;   TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
;   FILE REFERENCE: 529282000700
;   CURRENT APPLICATION NUMBER: US/10/174,695
;   CURRENT FILING DATE: 2002-06-18
;   PRIOR APPLICATION NUMBER: PCT/AU00/01588
;   PRIOR FILING DATE: 2000-12-21
;   PRIOR APPLICATION NUMBER: AU PQ 4859
;   PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

Query Match   30.0%; Score 791.5; DB 15; Length 419;
Best Local Similarity   39.9%; Pred. No. 8e-62;
Matches   174; Conservative 61; Mismatches 100; Indels 101; Gaps 14;

Qy   58 AGMARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYK 115
             :|:||||||||||||||||||||||||||||||:|||: : || || ||||||||:|
Db   1 SGQAEIVLEAHDVWNDGSGYQILLDADHDQYGVIPSDTHTLW-PNCSVPANLFAPPEYT 59

Qy   116 VPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAG 174
             |||||| ::|| ||:|||||||| |||||||:::: |||| : :|:|| :||| ||||
Db   60 VPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAG 119

Qy   175 KTYHFTVQRQGPGDAASVV------------------------------------- 193
             |:||||::||::||||||
Db   120 KKYHFLMKKMGSGDGTELITISEGGGSDYTYTVVRDGTKIKEGLTATTFEEDGVAAGNHEY 179

Qy   194 -------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------------------ 226
                    |||||||||||||||||||||||||||:||||
Db   180 CVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNP 239

Qy   227 --DKRTYVLNESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTAS 284
               ::||||||||||||||||||||||||||| :||||||||||||||||
Db   240 NPNPGTTLSESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GL 296

Qy   285 GGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKL 343
             :: |:|:|||||||| :|||||||||||||||| |||||||||||||||||||||||||
Db   297 GGIGV-LITPDNYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 355

Qy   344 LEETLGSDKPAPMNLVKSEGVKLP----------APYQERTIDLSAYAGGQVYLAFRHFNST 395
             |||| :: :||:||| |||||||||||||||
Db   356 LEET--------------ITAKGVRSPEAIRGRIOGTWRQKTVDLP--AGTK-YVAFRHFQST 401

Qy   396 GIFRLYLDDVAVSGEG 411
             :|:||| :|:| :
Db   402 DMFYIDLDEVEIKANG 417

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-5

Query Match   25.1%; Score 662.5; DB 15; Length 419;
Best Local Similarity   36.0%; Pred. No. 2.7e-50;
Matches   157; Conservative 60; Mismatches 118; Indels 101; Gaps 15;

Qy   58 AGMARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKV 116
             :|:|:||||||||:|||||||||||||||||||||| : || ||:|||||||| |||
Db   1 ANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLI 59

Qy   117 PVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVV 171
             | ||| |||||||||||||:||||||:|||||||||| || : :|| :||||||
Db   60 PANADPVTTQNIIVTGQGEVVIPGGVIDYCITNPEPASGKMWIAGDGGNQPARYDDFTF 119

Qy   172 EAGKTYHFTVQRQGPGDAASVV----------------------------------- 193
             ::||:|||||:||:||||||
Db   120 EAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGN 179

Qy   194 ------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP-------------- 224
                         |||||||||||||||||||||||||:||||
Db   180 HEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPN 239

Qy   225 ASDKRTYVLNESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTAS 284
             :: :|||||||||||||||||||||||||| :||||||||||||||||
Db   240 PNPNPGTTLSESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GL 296

Qy   285 GGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKL 343
             :: |:|:|||||||| :|||||||||||||||| ::: :||||||||||||||||||||
Db   297 GGIGV-LITPDNYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 355

Qy   344 LEETLGSDKPAPMNLVKSEGVKLP----------APYQERTIDLSAYAGGQVYLAFRHFNST 395
             |||| :::|| ||: |:||||||||||||||
Db   356 LEET--------------ITAKGVRSPKAIRGRIOGTWRQKTVDLP--AGTK-YVAFRHFQST 401

Qy   396 GIFRLYLDDVAVSGEG 411
             |||
Db   402 DMFYIDLDEVEIKANG 417

RESULT 5
US-10-174-695-6
; Sequence 6, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-6

Query Match   17.3%; Score 457.5; DB 15; Length 231;
Best Local Similarity   27.8%; Pred. No. 2.2e-32;
Matches   118; Conservative 22; Mismatches 63; Indels 221; Gaps   6;

Qy   80 LWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPNADASFSPTNFVLDGTASAD 138
             ||||||:|:: || ||:|||||||| |||| ||| |||||:|:||||||||||
Db   2 LLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVTTQNIIVTGQGEVV 60

Qy   139 IPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTVQRQGPGDAASVV 193
             ||:||:|||||||||| || : :|| :||||||::||:|||||:||:||||||
Db   61 IPGGVIDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDME 120

Qy   194 VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASADKRTYVLNESFDTQTLPNGWTMIDADG 253
             |||||||||||||||||||||||||:||||
Db   121 VE-- 122

```
Qy     254  DGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYW  313

Db     123  ------------------------------------------------------------  122

Qy     314  VSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDKPAPMNLVKSEGVKLPAPYQER  373

Db     123  ------------------------------------------DDSPA------------  127
                                                      ||||

Qy     374  TIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTVRDNVVIAQNLAA  433
                                                      ||||||  |  ::  |

Db     128  -------------------------------SYTYTVRDGTKIKEGLTA  146

Qy     434  TTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQKVTLKW  493
                :  :: |   ||||||||||||||||||:|||||:|||||:|||||||:|:|||||

Db     147  TTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKW  206

Qy     494  DAPN  497
            ||||
Db     207  DAPN  210
```

RESULT 6
US-10-174-695-4
; Sequence 4, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
;　APPLICANT:　Reynolds, Eric Charles
;　APPLICANT:　Slakeski, Nada
;　APPLICANT:　Chen, Chao Guang
;　APPLICANT:　Barr, Ian George
;　TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
;　FILE REFERENCE: 52928000700
;　CURRENT APPLICATION NUMBER: US/10/174,695
;　CURRENT FILING DATE:　2002-06-18
;　PRIOR APPLICATION NUMBER: PCT/AU00/01588
;　PRIOR FILING DATE:　2000-12-21
;　PRIOR APPLICATION NUMBER: AU PQ 4859
;　PRIOR FILING DATE: 1999-12-24
;　NUMBER OF SEQ ID NOS: 8
;　SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;　LENGTH: 196
;　TYPE: PRT
;　ORGANISM: Porphyromonas gingivalis
US-10-174-695-4

Query Match　　　7.1%;　Score 188.5; DB 15; Length 196;
Best Local Similarity　33.5%;　Pred. No. 1.8e-08;
Matches　59;　Conservative　8; Mismatches　50; Indels　59; Gaps　5;

```
Qy     194  VTGEGGNEFAPVQNLQWSVSGQTVLTWQAPAS------------DKRTYVLNRSFDTQT  241
            ||:||||||||||||  | |  |  | ||  |            |:||:||:::|:

Db       6  VTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESFE-NG  64

Qy     242  LPNGWTMIDADGDGHNWL----------------------STINVYNTATHTGDG  274
            :|  ||||||||||||||                      ||||:|
Db      65  IPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCVYLDNSAKIDRNQEINVYNTAEY----  120

Qy     275  AMFSKSWTASGGAKIDLSP--DNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGV  327
                                 ||  :: |   ||   :    |
Db     121  ------------AKTNNAPIKVVGYADEKTGTAAYNMKLSERRAKAVAKMLEKYGV  164
```

RESULT 7
US-10-387-977-27
; Sequence 27, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
;　APPLICANT:　Reynolds, Eric Charles
;　APPLICANT:　O'Brien-Simpson, Neil Martin
;　APPLICANT:　Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE

; TITLE OF INVENTION:　DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION:　PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200030l
;　CURRENT APPLICATION NUMBER: US/10/387,977
;　CURRENT FILING DATE:　2003-07-18
;　PRIOR APPLICATION NUMBER: US 09/423,056
;　PRIOR FILING DATE: 2000-03-22
;　PRIOR APPLICATION NUMBER: PCT/AU98/00311
;　PRIOR FILING DATE: 1998-04-30
;　PRIOR APPLICATION NUMBER: AU PO 6528
;　PRIOR FILING DATE: 1997-04-30
;　NUMBER OF SEQ ID NOS: 105
;　SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;　LENGTH: 29
;　TYPE: PRT
;　ORGANISM: Porphyromonas gingivalis
US-10-387-977-27

Query Match　　　5.7%;　Score 150;　DB 15;　Length 29;
Best Local Similarity　100.0%;　Pred. No. 2.7e-06;
Matches　29;　Conservative　0; Mismatches　0; Indels　0; Gaps　0;

```
Qy     415  DYTYTVYRDNVVIAQNLAATTFNQENVAP  443
            |||||||||||||||||||||||||||||
Db       1  DYTYTVYRDNVVIAQNLAATTFNQENVAP  29
```

RESULT 8
US-10-282-122A-51483
; Sequence 51483, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;　APPLICANT:　Wang, Liangsu
;　APPLICANT:　Zamudio, Carlos
;　APPLICANT:　Malone, Cheryl
;　APPLICANT:　Haselbeck, Robert
;　APPLICANT:　Ohlsen, Kari
;　APPLICANT:　Zyskind, Judith
;　APPLICANT:　Wall, Daniel
;　APPLICANT:　Trawick, John
;　APPLICANT:　Carr, Grant
;　APPLICANT:　Yamamoto, Robert
;　APPLICANT:　Forsyth, R.
;　APPLICANT:　Xu, H.
;　TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;　FILE REFERENCE: ELITRA.034A
;　CURRENT APPLICATION NUMBER: US/10/282,122A
;　CURRENT FILING DATE:　2003-02-20
;　PRIOR APPLICATION NUMBER: 60/191,078
;　PRIOR FILING DATE: 2000-03-21
;　PRIOR APPLICATION NUMBER: 60/206,848
;　PRIOR FILING DATE: 2000-05-23
;　PRIOR APPLICATION NUMBER: 60/207,727
;　PRIOR FILING DATE: 2000-05-26
;　PRIOR APPLICATION NUMBER: 60/230,335
;　PRIOR FILING DATE: 2000-09-06
;　PRIOR APPLICATION NUMBER: 60/230,347
;　PRIOR FILING DATE: 2000-09-09
;　PRIOR APPLICATION NUMBER: 60/242,578
;　PRIOR FILING DATE: 2000-10-23
;　PRIOR APPLICATION NUMBER: 60/253,625
;　PRIOR FILING DATE: 2000-11-27
;　PRIOR APPLICATION NUMBER: 60/257,931
;　PRIOR FILING DATE: 2000-12-22
;　PRIOR APPLICATION NUMBER: 60/267,636
;　PRIOR FILING DATE: 2001-02-09
;　PRIOR APPLICATION NUMBER: 60/269,308
;　PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51483

```
; LENGTH: 1483;
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51483

Query Match         5.7%;   Score 150; DB 12;  Length 1483;
Best Local Similarity  20.2%;  Pred. No. 0.0011;
Matches  127;  Conservative  81;  Mismatches 213;  Indels 208;  Gaps  29;

Qy   27 GGPKTAP-------SVTHQAVQKGIRTSKVKDLRDPIPAGMARIILEAHDVWEDGTGY-  77
Db  522 GTPVVTPSQINVEQGSATDQPVKIDLNGNTLKDVVDQ---SGKTLVQGTDYTVTDTGIT  577

Qy   78 ----QMLWDADHNQYGAS------IPEESFWFANGTIPAGLYDPPEYKVPV          118
Db  578 LSQSYLAGLALGQYTLTLDFNGGGASQTITINVVKNETVKLSVGTVSGNPGD-TVKVPV  635

Qy  119 N------------ADASFSPTNF-VLDGTASADIPAGT--YDYVIINPNPGIIYIV---  159
Db  636 TISQVSTPVGLICMDISYDASKFTVKDVLPNTDLVKDTDNYSFIVNTSTPGKISITFTDP 695

Qy  160 ------GEGVSKGNDYVVEAGKT--YHFTVQRQGPGDAASVVVTGEGGNEFAPVQNL    208
Db  696 TLANYPISVDGILAYLDFIINSNATAGDSALTV------DPATILVADEND------KDI 743

Qy  209 QWSVSGQTVTLTWQAPASDKRTVLNESFDT--QTLP------NGWTMIDA-DGDGH    256
Db  744 KDAASNGKITVTGSAPV--VQSSVVNTSSVTYDQNAPQDQAVSITFNGNTVKDVKDASG- 800

Qy  257 NWLSTINVYNTATHTGDGAMFSKSWTASGGA------KIDLSPDN------YLVTPKVTVPE 306
Db  801 ---NTLKAGSDYTATSDGITLSQSYLATLAAGTYYTIDFSAGNAGTFTVVVKGKTVVGS  857

Qy  307 NGKLSYWVSS------QVPWT---NEHYGVFLSTTGNEAANFTIK------LLEET-   347
Db  858 ATTLAVGTVSGKAGDTVKVPVTISKVTTPVGLICAEIDYDASKFTVKDVLPNTDLVKDTD 917

Qy  348 ----LGSDKPAPMNL-----------VKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRH 391
Db  918 NYSFIVNTSTPGKISITFTDPTLANYPISADGI--LAYLDFIINSNATAGDSALTV---  971

Qy  392 FNSTGIFRLYLDDVAVSGEGSSNDYT-----------                       417
Db  972 -NPSGFIIADENDKDIQDAASNGKITVTGSTPVAENSVVNTSSVTYDQNAPQDQAVSITL 1030

Qy  418 ------------YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKY-T         455
Db 1031 NGNTITDVKDASGNTLKAGSDYTVTSDGITLSQSYLAT----LAAGTYTIVDFSAGN   1084

Qy  456 AGVSPKVCKDVTVEGSNEFAHVQNLTGSA                               484
Db 1085 AGTFTVVVKAKTVVSSATTLAVGTVSGKA                               1113
```

RESULT 9
US-10-229-066-3
```
; Sequence 3, Application US/10229066
; Publication No. US2003015637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-3

Query Match         5.6%;   Score 148; DB 14;  Length 42;
Best Local Similarity  64.1%;  Pred. No. 7.3e-06;
Matches  25;  Conservative  7;  Mismatches  7;  Indels  0;  Gaps  0;

Qy   58 AGMARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEES 96
Db    1 SGQAEIVLEAHDVWNDGSGYQILLDADHDQYGQVIPSDT 39
```

RESULT 10
US-10-387-977-101
```
; Sequence 101, Application US/10387977
; Publication No. US2004000S276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION:   DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION:   PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-101

Query Match         5.5%;   Score 145; DB 15;  Length 509;
Best Local Similarity  19.1%;  Pred. No. 0.00061;
Matches  102;  Conservative  65;  Mismatches 155;  Indels 212;  Gaps  22;

Qy  114 YKVPVN---ADASF------SPTNFVLD------GTASADIPAGTY----D        145
Db   10 YNTPVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYND 69

Qy  146 YVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFTVQRQGPGDAASVVVTGEGGNEFAPV 205
Db   70 GLAASAAPVFLALVGD------------------------------TDVISGEKGKKTKKV 100

Qy  206 QNLQWS-VSG---QTVTLTWQAPASDKRTYVLNE--SPDTQTLPN----GWTMIDADGD  254
Db  101 TDLYSAVDGDYFPEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGAD  160

Qy  255 GHNWLSTI-----------------NVYN-------TATHTGDGAMFSKSWTASGG    286
Db  161 -YSWNSQVGQPTIKYGMQYYNQEHGYTDVYNYLKAPYTGCYSHLNTGVSFA-NYTAHGS  218

Qy  287 A------------------------KIDLSPDNYLVTPK------VTVPENGKLSYWV   314
Db  219 ETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRVKEKGAYAYIG 278

Qy  315 SSQVPWTNEHY------GVFLSTTGNEAANFTIKLLEET----------           347
Db  279 SSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH 338

Qy  348 ----LGSDKPAPMN-LVKSEGVKLPA--PYQERTIDLSAYAGQ               383
```

```
Db   339 AGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNTYTLPASLPQNQASYSIQASAGS 398
               ||              :|:      |           |:   :   |||
Qy   384 QVYLAFRHFNSTGIFRLYLIDDVAVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAP 443
               |:|:|          |:|||| :   ::
Db   399 Y----------------VAISKDG------VLYGTGVANASGVATVSMTKQITEN 431

Qy   444 GQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQKVTLKWDAPN 497
               ||||:|:|:||     ||||||:||::|
Db   432 GNY----DVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPS 481


RESULT 11
US-10-387-977-6
; Sequence 6, Application US/10387977
; Publication No. US20040005276A1
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR APPLICATION FILING DATE: 1998-04-30
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200030l
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-6

  Query Match             5.5%;  Score 144;  DB 15;  Length 31;
  Best Local Similarity  96.6%;  Pred. No. 1e-05;
  Matches    28;  Conservative   0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy   469 EGSNEFAHVQNLTGSAVGQKVTLKWDAPN 497
               |||||||||||||||||||||||||||||
Db     1 EGSNEFAPVQNLTGSAVGQKVTLKWDAPN 29


RESULT 12
US-10-282-122A-62947
; Sequence 62947, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: ZySkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

```
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62947
; LENGTH: 1946
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-282-122A-62947

  Query Match              5.4%;  Score 143;  DB 12;  Length 1946;
  Best Local Similarity   18.4%;  Pred. No. 0.0071;
  Matches   110;  Conservative   63;  Mismatches  214;  Indels  210;  Gaps  23;

Qy    22 TAAAQGPKTAPSVTHQAVQKGIRTSKVKDLRD---PIPAGMARILEAHDVWEDGTGY 77
             :       |  ::    |   |:       ::           :    : |||
Db   417 TATLLAGGLTVKDIHNKQIQVSADGIKFTDMNGGTSAPVAAGTTRIT----RDTIGF 469

Qy    78 QMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNADASFSPTNFVLDG-TAS 136
                    :  |   |        ::     |||  |  | |: | ::
Db   470 A---------KEGGTVDDNSPYLDNAKLKVG-------NVEINKDTGINAGNQKITGLTDG 514

Qy   137 ADIPAGTYDYVIINPNPGIYIVGEGVSKGNDYVVEAGKTYHFTVQRQGPGDAASVVVTG 196
                                      |     ||:||         :::
Db   515 ADDADAVTIKQLKNAKPNL--TPGNGIEINN------------------TNSLLVDA 551

Qy   197 EGGNEPAPVQNLQWSVSGQTVLTWQAPA----------SDKRTY-----------VLNESFD 238
             |||                  :      |: :|||  ::            :
Db   552 ANGNTVPP----SYTIGVKTTLTTNNGPSDSKFAVKDSNTHNNSLVTAKDLAGYLNEVNR 607

Qy   239 T-QTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
                                    :::                 |:|
Db   608 TADSALQSFTVKGGDAASNNSITLDKTSNSLTVAGDNGITVKTDTTQKKVTVGIDQANGL 667

Qy   298 VTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETL------- 348
             ||:|:   ||                 |:: |     ||
Db   668 TTPKLTVGSNSKTQLVIEQMV----------------IGNDTKNI-IKGLSSTLTDITNTDNT 713

Qy   349 ------------------------------GSDKPA-----------PMNLVKSEGVKLPAPYQ 371
                                                       |:    |
Db   714 HTTEQNDAQGKESNAASIKDVLNAGENLQGNGKPVDFVSYDTVNFANGDGTTATVTY- 772

Qy   372 ERTIDLSAYAGQQVYLIAFRHFN------STGIFRLYLDDVAV----SGEGSSNDYTYTVY 421
             |:                                |: |::
Db   773 -----------NYNEGKPTSTVVYNVNVDDTTIHLANNNGNKKLSVKTTTLT 813

Qy   422 RDNVVIAQNLAATTF-----NQENVAPGQYNYCVEVKYTAGVSP-----KVCKD- 465
                                 |    ::       |   : :  ||
Db   814 KTN---GANGNATKFSANNGDALVNAKGIADNLNILAEEIHTTKGTADTALQTFKVKKDG 870

Qy   466 -----------------------VTVEGSNEFA---------------HVQNLTGSAVGQKVTLKWD 494
                                     ::|   : |  |         ||||||||||:
Db   871 ATNDDDTITVGKNAVDTLTFKGENGLTVATKKDGTVTFGINTQSGLKAGDNTTLNKD 927


RESULT 13
US-10-387-977-9
```

```
;   Sequence 9, Application US/10387977
;   Publication No. US20040005276A1
;   GENERAL INFORMATION:
;    APPLICANT: Reynolds, Eric Charles
;    APPLICANT: O'Brien-Simpson, Neil Martin
;    APPLICANT: Slakeski, Nada
;    TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;    TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;    TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;    FILE REFERENCE: 52928200 0301
;    CURRENT APPLICATION NUMBER: US/10/387,977
;    CURRENT FILING DATE: 2003-07-18
;    PRIOR APPLICATION NUMBER: US 09/423,056
;    PRIOR FILING DATE: 2000-03-22
;    PRIOR APPLICATION NUMBER: PCT/AU98/00311
;    PRIOR FILING DATE: 1998-04-30
;    PRIOR APPLICATION NUMBER: AU PO 6528
;    PRIOR FILING DATE: 1997-04-30
;    NUMBER OF SEQ ID NOS: 105
;    SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 9
;    LENGTH: 31
;    TYPE: PRT
;    ORGANISM: Porphyromonas gingivalis
US-10-387-977-9

Query Match         5.3%;   Score 141;   DB 15;   Length 31;
Best Local Similarity   93.1%;   Pred. No. 1.9e-05;
Matches    27;   Conservative    1;   Mismatches   1;   Indels   0;   Gaps   0;

Qy     469  EGSNEFAHVQNLTGSAVGQKVTLKWDAPN 497
                ||||||| ||||||:|||||||||||||||
Db       1  EGSNEFAPVQNLTGSSVGQKVTLKWDAPN 29


RESULT 14
US-10-387-977-15
;   Sequence 15, Application US/10387977
;   Publication No. US20040005276A1
;   GENERAL INFORMATION:
;    APPLICANT: Reynolds, Eric Charles
;    APPLICANT: O'Brien-Simpson, Neil Martin
;    APPLICANT: Slakeski, Nada
;    TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;    TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;    TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;    FILE REFERENCE: 52928200 0301
;    CURRENT APPLICATION NUMBER: US/10/387,977
;    CURRENT FILING DATE: 2003-07-18
;    PRIOR APPLICATION NUMBER: US 09/423,056
;    PRIOR FILING DATE: 2000-03-22
;    PRIOR APPLICATION NUMBER: PCT/AU98/00311
;    PRIOR FILING DATE: 1998-04-30
;    PRIOR APPLICATION NUMBER: AU PO 6528
;    PRIOR FILING DATE: 1997-04-30
;    NUMBER OF SEQ ID NOS: 105
;    SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 15
;    LENGTH: 26
;    TYPE: PRT
;    ORGANISM: Porphyromonas gingivalis
US-10-387-977-15

Query Match         5.3%;   Score 139;   DB 15;   Length 26;
Best Local Similarity  100.0%;   Pred. No. 2.2e-05;
Matches    26;   Conservative    0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     202  FAPVQNLQWSVSGQTVTLTWQAPASD 227
                ||||||||||||||||||||||||||
Db       1  FAPVQNLQWSVSGQTVTLTWQAPASD 26


RESULT 15
```

```
US-10-282-122A-66335
;   Sequence 66335, Application US2004002912 9A1
;   Publication No. US2004002912 9A1
;   GENERAL INFORMATION:
;    APPLICANT: Wang, Liangsu
;    APPLICANT: Zamudio, Carlos
;    APPLICANT: Malone, Cheryl
;    APPLICANT: Haselbeck, Robert
;    APPLICANT: Ohlsen, Kari
;    APPLICANT: Zyskind, Judith
;    APPLICANT: Wall, Daniel
;    APPLICANT: Trawick, John
;    APPLICANT: Carr, Grant
;    APPLICANT: Yamamoto, Robert
;    APPLICANT: Forsyth, R.
;    APPLICANT: Xu, H.
;    TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;    FILE REFERENCE: ELITRA.034A
;    CURRENT APPLICATION NUMBER: US/10/282,122A
;    CURRENT FILING DATE: 2003-02-20
;    PRIOR APPLICATION NUMBER: 60/191,078
;    PRIOR FILING DATE: 2000-03-21
;    PRIOR APPLICATION NUMBER: 60/206,848
;    PRIOR FILING DATE: 2000-05-23
;    PRIOR APPLICATION NUMBER: 60/207,727
;    PRIOR FILING DATE: 2000-05-26
;    PRIOR APPLICATION NUMBER: 60/230,335
;    PRIOR FILING DATE: 2000-09-06
;    PRIOR APPLICATION NUMBER: 60/230,347
;    PRIOR FILING DATE: 2000-09-09
;    PRIOR APPLICATION NUMBER: 60/242,578
;    PRIOR FILING DATE: 2000-10-23
;    PRIOR APPLICATION NUMBER: 60/253,625
;    PRIOR FILING DATE: 2000-11-27
;    PRIOR APPLICATION NUMBER: 60/257,931
;    PRIOR FILING DATE: 2000-12-22
;    PRIOR APPLICATION NUMBER: 60/267,636
;    PRIOR FILING DATE: 2001-02-09
;    PRIOR APPLICATION NUMBER: 60/269,308
;    PRIOR FILING DATE: 2001-02-16
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;    NUMBER OF SEQ ID NOS: 78614
;    SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 66335
;    LENGTH: 2468
;    TYPE: PRT
;    ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66335

Query Match         5.2%;   Score 138.5;   DB 12;   Length 2468;
Best Local Similarity   23.7%;   Pred. No. 0.026;
Matches   107;   Conservative   47;   Mismatches  166;   Indels  131;   Gaps   23;

Qy     103  TIPAGLYDFPEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEG 162
                |||||||                                            ::  ||||||
Db     319  TDPAG-------NNSTPVTVEAP------DTTA----PAPATD-------VQVAPDG 351

Qy     163  VS-KGNDYVVEAGKTYHFTVQRQGPGDAASVVVTGEGNEFAPVQNLQWSVSGQTVTLTW 221
                :  :||::          ||||        ::||::|
Db     352  SSVTGN--AEPGAT--VGVDTDGDGQPDTTVVVGPGGSFEVPLN--PPLTNGETVTVIV 404

Qy     222  QAPASDKRTYVLNESFDTQTLPN------------------------GWTMIDADGDGH---------- 256
                                           ||::|:
Db     405  TDPAGNSSTPVTAEAPDFPDAPQVNASNGSVLSGTAEAGVTIVITDGNGNPIGQTSADAN 464

Qy     257  -NWLST------------INVY------NTATH-TGDGAM-----------FSKSWTASG 285
                  ||                :||::                ||:|
Db     465  GNWSFTPGSQLPDGTVNVVARDAAGNSSPATSIITVDGVAPNAPVEPSNGSELSGTAEP 524

Qy     286  GAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLE 345
                :::|||::||:||||||||||:|||:||:                |:|:|  :|||
Db     525  GSSVTLTDGNGNPIGQTTADANGNWSFTPSTPLP-----DGTVNVVARDAAGNSSPPAS 579
```

```
Qy  346  ETLGSDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDV 405
              |:.:||    ||     ||  |:.:|   |        |:.:||
Db  580  VTVDAVAPATPTVDPSNGT--------TLSGTAEPGSSVTLTDGNGNPIG--------- 621

Qy  406  AVSGEGSSNDYTYT----VYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPK 461
           |:.:||| .:|:|         :|  |          :||     |  :|
Db  622  QVTADGSGN-WTFTPSTPLPNGTVV---NATATDPSGNASSPAS----VTVDAVAPATPV 673

Qy  462  V--CKDVTVEGSNEFAHVQNLT---GSAVGQ 487
          |   |.|: |   ||  :|| :.:||
Db  674  VNPSNGTTLSGTAEPPGATVTLTDGNGNPIGQ 704
```

Search completed: May 18, 2004, 11:38:27
Job time : 22.8443 secs

```
              GenCore version 5.1.6
         Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May 18, 2004, 11:21:33 ; Search time 87.0274 Seconds
                                      (without alignments)
                                      5623.200 Million cell updates/sec

Title:       US-08-353-485-10
Perfect score: 9179
Sequence:    1 MRKLLLIASLLGVGLYAQ.........HYAVMVVVDGKSYVEKLAVK 1732

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:    1586107 seqs, 28254705 residues

Total number of hits satisfying chosen parameters:  ,  1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_29Jan04:*
             1: geneseqp1980s:*
             2: geneseqp1990s:*
             3: geneseqp2000s:*
             4: geneseqp2001s:*
             5: geneseqp2002s:*
             6: geneseqp2003as:*
             7: geneseqp2003bs:*
             8: geneseqp2004s:*

                        SUMMARIES
```

```
                %
Result          Query
No.    Score    Match  Length  DB  ID        Description
------------------------------------------------------------------------------
  1    9179     100.0  1732    2   AAR96029   Aar96029 P. gingiv
  2    9179     100.0  1732    2   AAW24787   Aaw24787 PrtK anti
  3    9179     100.0  1732    2   AAW69487   Aaw69487 Haemagglu
  4    6151.5   67.0   1358    2   AAR96032   Aar96032 P. gingiv
  5    6151.5   67.0   1358    2   AAW69494   Aaw69494 Haemagglu
  6    4903     53.4   970     2   AAR72458   Aar72458 Porphyrom
  7    4673.5   50.9   1087    2   AAR96028   Aar96028 P. gingiv
  8    4673.5   50.9   1087    2   AAW69486   Aaw69486 Haemagglu
  9    4272     46.5   1706    2   AAW24786   Aaw24786 PrtR anti
 10    4270     46.5   1704    3   AAY67396   Aay67396 Arg-gingi
 11    4270     46.5   1704    4   AAU08938   Aau08938 P. gingiv
 12    4264     46.5   1704    2   AAR70188   Aar70188 Arg-gingi
 13    4263     46.4   1704    2   AAW34843   Aaw34843 Arg-gingi
 14    4245     46.2   1687    2   AAW69495   Aaw69495 Haemagglu
 15    4233     46.1   1687    2   AAR96033   Aar96033 P. gingiv
 16    4217     45.9   2628    2   AAR96030   Aar96030 P. gingiv
 17    4217     45.9   2628    2   AAW69488   Aaw69488 Haemagglu
 18    2618.5   28.5   506     2   AAW83126   Aaw83126 PrtIIK48
 19    2565     27.9   509     2   AAW34847   Aaw34847 Lys-gingi
 20    2254     24.6   419     4   AAU03574   Aau03574 P. gingiv
 21    2034.5   22.2   456     2   AAR96023   Aar96023 P. gingiv
 22    2034.5   22.2   456     2   AAR96022   Aar96022 P. gingiv
 23    2034.5   22.2   456     2   AAW69491   Aaw69491 Haemagglu
 24    2034.5   22.2   456     2   AAW69490   Aaw69490 Haemagglu
 25    2025.5   22.1   439     2   AAR96024   Aar96024 P. gingiv
 26    2025.5   22.1   439     2   AAW69492   Aaw69492 Haemagglu
 27    2011.5   21.9   450     2   AAR96021   Aar96021 P. gingiv
 28    2011.5   21.9   450     2   AAW69489   Aaw69489 Haemagglu
 29    1854     20.2   364     2   AAR96031   Aar96031 P. gingiv
 30    1854     20.2   364     2   AAW69493   Aaw69493 Protease
 31    1776     19.3   419     4   AAU03572   Aau03572 P. gingiv
 32    1241     13.5   231     4   AAU03575   Aau03575 P. gingiv
 33    1221.5   13.3   991     2   AAR77313   Aar77313 Porphyrom
 34    980.5    10.7   921     2   AAR34522   Aay34522 Porphyrom
 35    980.5    10.7   922     2   AAY34521   Aay34521 Porphyrom
 36    980.5    10.7   925     2   AAY34520   Aay34520 Porphyrym
 37    980.5    10.7   938     2   AAY34392   Aay34392 Porphyrym
 38    812.5    8.9    497     2   AAR96025   Aar96025 P. gingiv
 39    812.5    8.9    497     2   AAW69483   Aaw69483 Haemagglu
 40    763      8.3    148     2   AAB14942   Aab14942 Porphyrom
 41    749      8.2    377     2   AAY34359   Aay34359 Porphyrom
 42    725      7.9    135     6   ABP55081   Abp55081 Porphyrom
 43    722      7.9    312     2   AAY34484   Aay34484 Porphyrym
 44    715      7.8    134     4   AAB49217   Aab49217 Peptide u
 45    698      7.6    737     2   AAR70186   Aar70186 Arg-gingi
```

ALIGNMENTS

```
RESULT 1
AAR96029
ID   AAR96029 standard; protein; 1732 AA.
XX
AC   AAR96029;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis porphypain.
XX
KW   Porphypain; haemagglutinin; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain W12.
XX
FH   Key          Location/Qualifiers
FT   Region       688..708
FT                /note= "Pro-Asn repeat region type 1"
FT   Region       887..952
FT                /note= "Pro-Asn repeat region type 2"
FT   Region       946..967
FT                /note= "Pro-Asn repeat region type 1"
FT   Region       985..1006
FT                /note= "Pro-Asn repeat region type 3"
FT   Region       1041..1100
FT                /note= "Pro-Asn repeat region type 4"
FT   Region       1341..1405
FT                /note= "Pro-Asn repeat region type 2"
FT   Region       1430..1451
FT                /note= "Pro-Asn repeat region type 3"
FT   Region       1488..1547
FT                /note= "Pro-Asn repeat region type 4"
FT   Region       1607..1650
FT                /note= "Pro-Asn repeat region type 2"
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;  95WO-US016108.
XX
PR   09-DEC-1994;  94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30653.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 76-81; 153pp; English.
XX
CC   P. gingivalis W12 cysteine protease, porphypain (AAR96029), was
CC   identified as the product of the prtP gene (AAT30653) isolated from P.
CC   gingivalis W12 genomic DNA. The porphypain shows homology to the
CC   haemagglutinins (see also AAR96026-28 and AAR96030-33) of P. gingivalis
CC   318. It can be obtd. from transformed host cells and used as a vaccine to
CC   protect humans or animals against periodontal disease. Expression in
CC   Salmonella cells allows prodn. of a live vaccine. The porphypain and
CC   haemagglutinins can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1732 AA;

Query Match             100.0%;   Score 9179;   DB 2;   Length 1732;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 1732;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy      1 MRKLLLIIAASLLGVGLYAQSAKIKLDAPTTRTTCNNNSFKQFDASFSFNEVELTKVETK 60
            |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      1 MRKLLLIIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK 60

Qy     61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL 120

Qy    121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180

Qy    181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VVANQLKVRNNIEIEVSFQGADEVATORLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240

Qy    241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300

Qy    301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360

Qy    361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNY 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNY 420

Qy    421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480

Qy    481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS 540

Qy    541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP 600

Qy    601 KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660

Qy    661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720

Qy    721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Qy    781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI 840

Qy    841 AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYYTVYRDGTKIKE 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYYTVYRDGTKIKE 900

Qy    901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960

Qy    961 TLKWDAPNGTPNPNPNENPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TLKWDAPNGTPNPNPNENPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020

Qy   1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASBHYAVYASS 1080

Qy   1081 TGNDASNFTNALLEETTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TGNDASNFTNALLEETTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140

Qy   1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200

Qy   1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA 1260

Qy   1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320

Qy   1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380

Qy   1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440

Qy   1441 SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP 1500

Qy   1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560

Qy   1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTTYIYRNNTQIAS 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTTYIYRNNTQIAS 1620

Qy   1621 GVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680

Qy   1681 TVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYAVMVVVDGKSYVEKLAVK 1732
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYAVMVVVDGKSYVEKLAVK 1732

RESULT 2
AAW24787
ID   AAW24787 standard; protein; 1732 AA.
XX
```

```
AC   AAW24787;
XX
DT   17-OCT-2003  (revised)
DT   25-NOV-1997  (first entry)
XX
DE   PrtK antigenic protein complex.
XX
KW   Periodontal disease; cell surface protein; thiol protease; endopeptidase;
KW   PrtK; PrtK48; PrtK39; PrtK15; PrtK44; haemagglutinin; adhesin; therapy;
KW   diagnosis; vaccine; antigen.
XX
OS   Porphyromonas gingivalis; strain W50.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..228
FT                   /label= Pro-pro_peptide
FT   Cleavage-site   228..229
FT   Protein         229..737
FT                   /label= PrtK48
FT                   /note= "48 kDa Lys-specific thiol protease"
FT   Cleavage-site   737..738
FT   Protein         738..1156
FT                   /label= PrtK39
FT                   /note= "39 kDa adhesin"
FT   Cleavage-site   1156..1157
FT   Protein         1157..1291
FT                   /label= PrtK15
FT                   /note= "15 kDa adhesin"
FT   Cleavage-site   1291..1292
FT   Protein         1292..1732
FT                   /label= PrtK44
FT                   /note= "44 kDa adhesin"
XX
PN   WO9716542-A1.
XX
PD   09-MAY-1997.
XX
PF   30-OCT-1996;  96WO-AU000673.
XX
PR   30-OCT-1995;  95AU-00006275.
XX
PA   (UYME ) UNIV MELBOURNE.
PA   (VICT-) VICTORIAN DAIRY IND AUTHORITY.
XX
PI   Reynolds EC,  Bhogal PS,  Slakeski N;
XX
DR   WPI; 1997-272112/24.
DR   N-PSDB; AAT78851.
XX
PT   New antigenic protein complex from Porphyromonas gingivalis - comprising
PT   Arg- and Lys- specific thiol endo-peptidase(s), used in the detection,
PT   prevention and treatment of periodontal disease.
XX
PS   Example 1; Fig 9b; 68pp; English.
XX
CC   A PrtR-PrtK cell surface protein of Porphyromonas ginivalis (PG)
CC   comprises a 300 kDa complex composed a 48 kDa lysine-specific thiol
CC   protease and 39, 15 and 44 kDa adhesins encoded by the prtK gene
CC   (AAT78851), and a 45 kDa arginine-specific thiol protease and 44, 15, 17
CC   and 27 kDa adhesins (see AAW24786) encoded by the prtR gene (AAT78850). A
CC   claimed antigenic complex comprises at least one multimeric protein
CC   complex of PrtR and PrtK each containing at least one adhesin domain, the
CC   complex having a mol.wt. of over 200 kDa, and preferably comprises all 9
CC   proteins of the PrtR-PrtK complex (see also AAW24780-85). It can be used
CC   in a claimed composition to elicit an immune response directed against
CC   PG, and in a claimed method of reducing the prospect of PG infection
CC   and/or severity of disease. Antibodies directed against the complex are
CC   claimed for use in treating PG infection. Unlike whole PG cells or other
CC   previously prepared antigens based on fimbriae or the capsule, the PrtR-
CC   PrtK complex or component parts are safe and effective antigens. (Updated
CC   on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1732 AA;
```

```
     Query Match             100.0%;  Score 9179;  DB 2;  Length 1732;
     Best Local Similarity   100.0%;  Pred. No. 0;
     Matches 1732;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1  MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELIKVETK   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELIKVETK   60

Qy     61  GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRKSFTEQVVSLNQYGSEKL  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRKSFTEQVVSLNQYGSEKL  120

Qy    121  MPHQPSMSKSDDPEKVPFVVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  MPHQPSMSKSDDPEKVPFVVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD  180

Qy    181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT  240

Qy    241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA  300

Qy    301  ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360

Qy    361  IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNY  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNY  420

Qy    421  LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA  480

Qy    481  QFDVVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  QFDVVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS  540

Qy    541  YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP  600

Qy    601  KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG  660

Qy    661  NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK  720

Qy    721  RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT  780

Qy    781  GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVIPGGVVDYCITNPEPASGKMWI  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVIPGGVVDYCITNPEPASGKMWI  840

Qy    841  AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTTVYRDGTKIKE  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTTVYRDGTKIKE  900

Qy    901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV  960

Qy    961  TLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  TLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020
```

```
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58874.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 69-84; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the prtP haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1732 AA;

Query Match               100.0%;  Score 9179;  DB 2;  Length 1732;
Best Local Similarity     100.0%;  Pred. No. 0;
Matches  1732;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy    1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTRTTCTNNSFKQFDASFSFNEVELTKVETK 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTRTTCTNNSFKQFDASFSFNEVELTKVETK 60

Qy   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL 120

Qy  121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180

Qy  181 VVANQLKVRNNIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 VVANQLKVRNNIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240

Qy  241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTNASIKAFIHKKYNDGLA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTNASIKAFIHKKYNDGLA 300

Qy  301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360

Qy  361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNY 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNY 420

Qy  421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDXYFLAIGNCCITA 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDXYFLAIGNCCITA 480

Qy  481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS 540

Qy  541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP 600

Qy  601 KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660

Qy  661 NYDVVITRSNYLPVIKQIQVGBPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 NYDVVITRSNYLPVIKQIQVGBPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720

Qy 1021 NSNGCVVSESFGLGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 NSNGCVVSESFGLGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080

Qy 1081 TGNDASNFTNALLEETTTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 TGNDASNFTNALLEETTTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140

Qy 1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200

Qy 1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260

Qy 1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320

Qy 1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380

Qy 1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440

Qy 1441 SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP 1500
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP 1500

Qy 1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560

Qy 1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIAS 1620
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIAS 1620

Qy 1621 GVTETTYRDPDLIATGFVTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 GVTETTYRDPDLIATGFVTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680

Qy 1681 TVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYAVMVVVDGKSYEKLAVK 1732
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 TVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYAVMVVVDGKSYEKLAVK 1732
```

```
RESULT 3
AAW69487
ID   AAW69487 standard; protein; 1732 AA.
XX
AC   AAW69487;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein prtP.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; prtP.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;  95US-00570311.
XX
PR   08-SEP-1988;  88US-00241640.
PR   25-JAN-1991;  91US-00647119.
PR   09-DEC-1994;  94US-00353485.
XX
```

```
XX
AC   AAR96032;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis hagD haemagglutinin.
XX
KW   Haemagglutinin; hagD; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain FDC381.
XX
PN   WO9617935-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;  95WO-US016108.
PR   09-DEC-1994;  94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A,  Tumwasorn S,  Lepine G,  Han N,  Lantz M,  Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30655.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 125-129; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagD (AAR96032) was identified as the
CC   product of the second open reading frame of the hagD gene (AAT30655)
CC   derived from P. gingivalis 318 genomic DNA. A first open reading frame
CC   coded for hagD protease (see also AAR96031). The protease and
CC   haemagglutinin can be obtd. from transformed host cells and used in
CC   vaccines to protect humans or animals against periodontal disease.
CC   Expression in Salmonella cells allows prodn. of live vaccines. The
CC   haemagglutinin and protease can also be used to detect the presence of
CC   anti-P. gingivalis antibodies and to raise monoclonal antibodies for
CC   diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1358 AA;

  Query Match             67.0%;  Score 6151.5;  DB 2;  Length 1358;
  Best Local Similarity   85.0%;  Pred. No. 0;
  Matches 1170;  Conservative 60;  Mismatches 118;  Indels 29;  Gaps 10;

Qy    366 MYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNYLKAPY 425
          ||||||||||||||| |||||||| |||  :|| ||||  |||:|| |||||||  |||||
Db      1 MYEKATMPDKSYLEKALLIAGADSYWNPKIGQQTIKYAVQYYYNQDHGYTDVYSYPKAPY 60

Qy    426 TGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYV 485
          |||||||||||.|||||||||||.||||.|.|:|:.|||||||||||||||||.|||||:
Db     61 TGCYSHLNTGVGFANYTAHGSETSWADPSLTATQVKALTNKDKYFLAIGNCCVTAQFDYP 120

Qy    486 QPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATF 545
          ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 QPCFGEVMTRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATF 180

Qy    546 LEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNTY 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNTY 240

Qy    606 TLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGYDVV 665
          ||||||||||||||||||||||||||||||||||||||||||||||:||||||||| |||
Db    241 TLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVNMTKQITENGNYDVV 300

Qy    666 ITRSNYLPVIKQIOVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAGGSREVKRIGDG 725
          ||||||||||||| .|||||||||||||||||||||||||:|||||||:..|||||||||
Db    301 ITRSNYLPVIKQIAGEPSPYQPVSNLTATTQGQKVTLKWDAPSAKKAEASREVKRIGDG 360

Qy    721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Qy    781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGGGEVVIPGGVYDYCITNPEPASGKMWI 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGGGEVVIPGGVYDYCITNPEPASGKMWI 840

Qy    841 AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900

Qy    901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960

Qy    961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020

Qy   1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080

Qy   1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140

Qy   1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200

Qy   1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA 1260

Qy   1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320

Qy   1321 NYILLDDIQFTMGGSPTPYDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 NYILLDDIQFTMGGSPTPYDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380

Qy   1381 AGVSPKKCCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDPENGIPA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AGVSPKKCCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDPENGIPA 1440

Qy   1441 SWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVSSASHINFEGQNPDNYLVTPELSLP 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 SWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVSSASHINFEGQNPDNYLVTPELSLP 1500

Qy   1501 GGGLTFWVCAQDANYASEHYAVVASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GGGLTFWVCAQDANYASEHYAVVASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560

Qy   1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIAS 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIAS 1620

Qy   1621 GVTETTYRDPDLATGFYTTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GVTETTYRDPDLATGFYTTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680

Qy   1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
```

RESULT 4
AAR96032
ID   AAR96032 standard; protein; 1358 AA.

```
XX
AC   AAW69494;
XX
DT   22-DEC-1998  (first entry)
XX
DE   Haemagglutinin protein hagD.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagD.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58880.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 145-158; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagD haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1358 AA;

Query Match             67.0%;  Score 6151.5;  DB 2;  Length 1358;
Best Local Similarity   85.0%;  Pred. No. 0;
Matches 1170;  Conservative 60;  Mismatches 118;  Indels 29;  Gaps 10;

Qy   366 MYEKATMPDKSYLEKVLLIAGADYSWNSQVSQPTIKYGMQYYNQEHGYTDVYNYLKAPY  425
Db     1 MYEKAIMPDKSYLEKALLIAGADSYWNPKIGQQTIKYAVQYYNQDHGYTDVYSYPKAPY   60

Qy   426 TGCYSHLNTGVSFANVTAHGSETAWADPLLITSQLKALTNKDKYFLAIGNCCITAQFDYV  485
Db    61 TGCYSHLNTGVGFANYTAHGSETSWADPSLTATQVKALTNKDKYFLAIGNCCVTAQFDYP  120

Qy   486 QPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYWSVGANAVFGVGVQPTFEGTSMGSYDATF  545
Db   121 QPCFGEVMTRVKEKGAYAYIGSSPNSYWGEDYWSVGANAVFGVGVQPTFEGTSMGSYDATF  180

Qy   546 LEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNTY  605
Db   181 LEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNTY  240

Qy   606 TLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVV  665
Db   241 TLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVNMTKQITENGNYDVV  300

Qy   666 ITRSNYLPVIKQIVQGEPSPYQPVSNLTATQGQKVTLKWEAPSAKKAEGSREVKRIGDG  725
Db   301 ITRSNYLPVIKQIQAGEPSPYQPVSNLTATTQGQKVTLKWDAPSAKKAEASREVKRIGDG  360
```

```
Qy   726 LFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASS  785
Db   361 LFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASS  420

Qy   786 NLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGG  845
Db   421 NLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGD  480

Qy   846 NQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTAT  905
Db   481 NQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTAT  540

Qy   906 TFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWD  965
Db   541 TFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTIKWD  600

Qy   966 APNGTPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGC  1025
Db   601 APNGTPNPNFNPNPGT-TTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGC  659

Qy  1026 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085
Db   660 VYSESFGLGGIGVLTPDNYLITPALDLLANGKLTFWVCAQDANYASEHYAVYASSTGNDA  719

Qy  1086 SNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDL 1145
Db   720 SNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDL  779

Qy  1146 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGSN  1205
Db   780 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTN  839

Qy  1206 VVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 1265
Db   840 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV  899

Qy  1266 VFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1325
Db   900 VFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYILL  959

Qy  1326 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1385
Db   960 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1019

Qy  1386 KKCVDVTVNSTQFNPVQNLTAEQAPNSMDALLKWNAPASKRABVLNEDFE-NGIPASWKT 1444
Db  1020 KVCVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSGKRGELLNEDFEGDAIPTGWTA 1077

Qy  1445 IDADGDGNNW---TTTPPPGG----SSFAGHNSAICVSSASHINFEGPQNPDNYLVTPEL 1497
Db  1078 LDADGDGNNWDITLNEFTRGERHVLSPLRASNVAISYSSLLQGQEYLPLITPENNFLITPKV 1137

Qy  1498 SLPGGGTLTFWVCAQD-ANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAI 1556
Db  1138 E--GAKKITYKVGSPGLPQWSHDHYALCISKSGTAAADF------EVIFEETMTYTQGGA 1189

Qy  1557 RGTRAQGTWTYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVIT-SGNAPSYTYIYRNN 1615
Db  1190 NLTREK-------DLPAGTKYVAFRHYNCTDVLGIMIDDVVIITGEGPSYTYTVYRDG 1241

Qy  1616 TQIASGVTETTYRDDPLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTV 1675
Db  1242 TKIQEGLTETTYRDAGMSAQSHEYCVEVKIAAGVSPKVCVDYIPDGVADVTAQKPYTLTV 1301

Qy  1676 VGKTIIVTCQGEAMIYDMNGRRLAAGRNTVVYTAQQGHYAVMVVVDGKSYVEKLAVK 1732
Db  1302 VGKTIIVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGYYAVMVVVDGKSYVEKLAIK 1358
```

```
RESULT 5
AAW69494 standard; protein; 1358 AA.
ID   AAW69494 standard; protein; 1358 AA.
```

DT   19-DEC-1995 (first entry)
DE   Porphyromonas gingivalis lysine-gingipain protein complex.
XX
KW   Lysine-gingipain protein complex; amidolytic; proteolytic;
KW   lysine specific proteinase; modulator identification; periodontitis;
XX   therapy monitoring.
OS   Porphyromonas gingivalis.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..228
FT   /label= sig_peptide
FT   Peptide   229..970
FT   /label= mat_peptide
XX
PN   WO9511298-A1.
XX
PD   27-APR-1995.
XX
PF   21-OCT-1994;   94WO-US012094.
XX
PR   21-OCT-1993;   93US-00141324.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,   Potempa JS,   Barr P,   Pavloff N,   Pike RN;
XX
DR   WPI; 1995-170220/22.
DR   N-PSDB; AAQ88141.
XX
PT   Lys-gingipain complex prepn with amidolytic and proteolytic specificity -
PT   for cleavage of an amide bond with Lysine contributing the carboxyl gp.
XX
PS   Claim 2; Page 53-57; 75pp; English.
XX
CC   AAQ88141 encodes AAR72458 the Porphyromonas gingivalis lysine-gingipain
CC   protein complex (LGPC). The LGPC has amidolytic and proteolytic
CC   specificity for an amide bond, where lysine contributes the carboxyl gp.,
CC   i.e. a lysine specific protease. LGPC can be used to identify agents that
CC   modulate the effect of LGPC on animals, and also for monitoring the
CC   exposure of an animal to LGPC. Such a method can be used to monitor the
CC   progress of a therapy designed to lessen the symptoms of periodontitis
XX
SQ   Sequence 970 AA;

Query Match   53.4%;   Score 4903;   DB 2;   Length 970;
Best Local Similarity 95.6%;   Pred. No. 1.7e-316;
Matches 927;   Conservative 21;   Mismatches 22;   Indels   0;   Gaps   0;

Qy   1 MRKLLLIIAASLLGVGLYAQSAKIKLDAPTRRTCTNNSFKQFDASFSFNEVELTKVETK   60
           ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   1 MRKLLLIIAASLLGVGLYAQNAKIKLDAPTTRTCTNNSFKQFDASFSFNEVELRKVETK   60

Qy   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVVSLNQYGSEKL 120
           |||||||||||||||||||||||||||||||||||| |||||||||||||| |||||||
Db   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVEARPVVRVKSFTEQVVCLNQYGSEKL 120

Qy   121 MPHQPSMSKSDDPEKVPFVVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 MPHQPSMSKSDDPEKIPFAYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180

Qy   181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDRGDLYNT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDRGDLYNT 240

Qy   241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300

Qy   301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360
           |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATAAPVFLALVGDTDVISGEKGKKTKKVTDLYYTAVDGDYFPEMYTFRMSASSPEELTNI 360

Db   Db
Qy

Db   361 LFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASS 420

Qy   786 NLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGG 845
           |||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 NLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGD 480

Qy   846 NQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTAT 905
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 NQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTAT 540

Qy   906 TFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWD 965
           |||||||:||||||||||||||||||||:||||||||||||||||||:||||||||||||
Db   541 TFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWD 600

Qy   966 APNGTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGC 1025
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 APNGTPNPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGC 659

Qy   1026 VYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085
           ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   660 VYSESFGLGGIGVLTPDNYLITPALDLLANGGKLTFWVCAQDANYASEHYAVYASSTGNDA 719

Qy   1086 SNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDL 1145
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   720 SNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDL 779

Qy   1146 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDCQGWLCLSSGQLDWLTAHGGSN 1205
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db   780 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGTN 839

Qy   1206 VVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 1265
           ||:||:|||||||||||||||||||||||||:|:|||||||||||||||||||||||||
Db   840 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 899

Qy   1266 VFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1325
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   900 VFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDIDYILL 959

Qy   1326 DDIQFTMGGSPTPPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1385
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   960 DDIQFTMGGSPTPPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1019

Qy   1386 KKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFE-NGIPASWKT 1444
           ||:|:||:|:|||||| :|| :||| ||:| ||| |: :||| :||:|| :|||||||
Db   1020 KVCVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSGKRGELLNEDFEGDAIPTGWTA 1077

Qy   1445 IDADGDGNNW--TTTPPGG-----SSFAGHNSAICVSSASHINFEGPQNPDNYLVTPEL 1497
           :||||||||| :||:| || ::|:|| :| ::|::|:||||| |||||||: ||:
Db   1078 LDADGDGNNWDITTLNEFTRGERHVLSPLRASNVAISYSSLLQGQEYLPLTPNNFLITPKV 1137

Qy   1498 SLPGGGTLTFWVCAQD-ANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAI 1556
           ::||| :||||||||| ||:|:||:|||:|: : |||::|:||::|| ||:|: : :||: :
Db   1138 E--GAKKITYKVGSPGLPQWSHDHYALCISKSGTAAADF-----EVIFEETMTYTQGGA 1189

Qy   1557 RGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLLDDVVIT-SGNAPSYTYTIYRNN 1615
           :||:| || ||::||||||||||||||: |||::||||||:|| |||||||||:|||:|:
Db   1190 NLTREK------DLPAGTKYVAFRHYNCTDVLGIMIDDVVITGEGEPSYTYTVYRDG   1241

Qy   1616 TQIASGVTETTYRPDDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTV 1675
           | |:| :|| :|||| ||||||| ||:|||| |||::|:|:|::|:||:|| |||||:||
Db   1242 TKIQEGLTETTYRDAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTV 1301

Qy   1676 VGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGHYAVMVVVDGKSYVEKLAVK 1732
           |||||||||||||||||||||||||||||:|||||||||:|:|||||||||||::
Db   1302 VGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGYYAVMVVVDGKSYVEKLAIK 1358

RESULT 6
AAR72458
ID   AAR72458 standard; protein; 970 AA.
XX
AC   AAR72458;
XX

```
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30652.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 65-68; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagD (AAR96028) was identified as the
CC   product of a gene (AAT30652) isolated from a P. gingivalis 318 genomic
CC   library. The haemagglutinin (see also AAR96032) can be obtd. from
CC   transformed host cells and used as a vaccine to protect humans or animals
CC   against periodontal disease. Expression in Salmonella cells allows prodn.
CC   of a live vaccine. The haemagglutinin can also be used to detect the
CC   presence of anti-P. gingivalis antibodies and to raise monoclonal
CC   antibodies for diagnostic appln. (Updated on 16-OCT-2003 to standardise
CC   OS field)
XX
SQ   Sequence 1087 AA;

Query Match             50.9%;  Score 4673.5;  DB 2;  Length 1087;
Best Local Similarity   81.4%;  Pred. No. 3.7e-301;
Matches  900;  Conservative   56;  Mismatches 120;  Indels   29;  Gaps   10;

Qy  638 GTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIKQIVGEPSPYQPVSNLTATTQ 697

Db    2 GTVVADPTVAAPVKMAKQIAENGNYDVVMTRSNYLPVINQIQAGEPSPYQPVNNLTAPPE 61

Qy  698 GQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGY 757

Db   62 GEEVALKWDTPSAKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGY 121

Qy  758 QFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLIVPANADPVVTTQNIIVTGQGEV 817

Db  122 QFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEV 181

Qy  818 VIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDM 877

Db  182 VIPGGVYDYCITNPEPASGKMWIAGDGDNQPARYDDFTFEAGKKYTFMRRAGMGDGTDM 241

Qy  878 EVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDV 937

Db  242 EVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDV 301

Qy  938 TVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASW 997

Db  302 TVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGT-TTLSESFENGIPASW 360

Qy  998 KTIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGK 1057

Db  361 KTIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLIITPALDLANGGK 420

Qy 1058 LTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETTAKGVRSPKAIRGRIQGTWR 1117

Db  421 LTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETTTAKGVRSPEAIRGRIQGTWR 480

Qy 1118 QKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTT 1177

Db  481 QKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTT 540

Qy 1178 IDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYY 1237

Db  541 IDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYY 600

Qy 1238 YAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWI 1297

Db  601 YAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWI 660

Qy 1298 ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLT 1357

Db  661 ERTVDLPAGTKYVAFRHYNCSDLDYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLT 720
```

```
Qy  361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVVNY 420

Db  361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQDHGYTDVVTY 420

Qy  421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480

Db  421 PKAPYTGCYSHLNTGVGFANYTVHGSEISWADPSVTATQVKALTNKNKYFLAIGNCCVTA 480

Qy  481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYIWSVGANAVFGVQPTFEGTSMGS 540

Db  481 QFDYPQPCFGEVMTRVKEKGAYAYIGSSPNSYWGEDYIWSVGANAVFGVQPTFEGTSMGS 540

Qy  541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP 600

Db  541 YDATFLEDSYNTVNSIMWAGNLAATHAENIGNVTHIGAHYYWEAYHVLGDGSVMPYRAMP 600

Qy  601 KTNYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660

Db  601 KTNYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVNMTKQITENG 660

Qy  661 NYDVVITRSNYLPVIKQIVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720

Db  661 NYDVVITRSNYLPVIKEIQAGEPSPYQPVSNLTATTQGQKVTLKWDAPSAKKAEGSREVK 720

Qy  721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Db  721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Qy  781 GTASSNLYSANFEYLIVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI 840

Db  781 GRASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITKPEPASGKMWI 840

Qy  841 AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900

Db  841 AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900

Qy  901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960

Db  901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKV 960

Qy  961 TLKWDAPNGT 970

Db  961 TLKWDAPNGT 970
```

```
RESULT 7
AAR96028
ID   AAR96028 standard; protein; 1087 AA.
XX
AC   AAR96028;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagD.
XX
KW   Haemagglutinin; hagD; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain FDC381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995; 95WO-US016108.
XX
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
```

```
Qy   1358  ETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAIL 1417
           ||||||||||||||||||||||||||||||||||:|::|: |:|
Db    721  ETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTIINPTQFNPVKNLKAQ--PDGGDVVL 778

Qy   1418  KWNAPASKRAEVLNEDFE-NGIPASWKTIDADGDGNNW--TTTPPGG-----SSFAGHN 1469
           ||:|| :|:|| |:||||||:|:: :|::  ::|
Db    779  KWEAPSGKRGELLNEDFEGDAIPTGWTALDADGDGNNWDITLNEFTRGERHVLSPLRASN 838

Qy   1470  SAICVSSASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQD-ANYASEHYAVYASST 1528
           ||:|:|:||| |:|:|| ::  :|::||:|:
Db    839  VAISYSSLLQGQEYLPLTPNNFLITPKVE--GAKKITYKVGSPGLPQWSHDHYALCISKS 896

Qy   1529  GNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTD 1588
           |||| :| :||| :|| |:: ::|
Db    897  GTAAADF------EVIFEETMTYTQGGANLTREK------DLPAGTKYVAFRHYNCTD 942

Qy   1589  FFWINLDDVVIT-SGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPN 1647
           ||||||||||||||:|:|
Db    943  VLGIMIDDVVITGEGEGPSYTYTVYRDGTKIQEGLTETTYRDAGMSAQSHEYCVEVKYAA 1002

Qy   1648  GESAIETATLNITSLADVTAQKPYTLTVVGKTITVVCQGEAMIYDMNGRRLAAGRNTVVY 1707
           |||||||:|:|
Db   1003  GVSPKVCVDYIPDGVADVTAQKPYTLITVVGKTITVVCQGEAMIYDMNGRRLAAGRNTVVY 1062

Qy   1708  TAQGGHYAVMVVVDGKSYVEKLAVK 1732
           |||||||||||||:|||||||:|
Db   1063  TAQGGYYAVMVVVDGKSYVEKLAIK 1087
```

RESULT 8

```
AAW69486
ID   AAW69486  standard; protein; 1087 AA.
XX
AC   AAW69486;
XX
DT   22-DEC-1998  (first entry)
XX
DE   Haemagglutinin protein hagD.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagD.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   1998-582627/49.
DR   N-PSDB; AAV58873.
XX
PR   08-SEP-1988;   95US-00570311.
PR   11-DEC-1995;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL )  UNIV FLORIDA.
PA   (UABR-)  UAB RES FOUND.
XX
PI   Patti JM,  Han N,  Lantz M,  Tumwasorn S,  Progulske-Fox A,  Lepine G;
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 57-64; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagD haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
```

SQ   Sequence 1087 AA;

```
Query Match             50.9%;  Score 4673.5;  DB 2;  Length 1087;
Best Local Similarity   81.4%;  Pred. No. 3.7e-301;
Matches  900; Conservative  56; Mismatches 120; Indels  29; Gaps  10;

Qy    638  GTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQ 697
           ||:: :: :|||:  :|:|:  || :||:|||:|||||:|||:||:||||||:||:||||
Db      2  GTVVADPTVAAPVKMAKQIAENGNYDVVMTRSNYLPVINQIQAGEPSPYQPVNNLTAPPE 61

Qy    698  GQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGY 757
           |  |:||||||:||| |: |||:|||||||||||||||:||||||||||:||||||||||
Db     62  GEEVALKWDTPSAKKABASREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGY 121

Qy    758  QFLLDADHNTFGSVIPAIGPLFTGTASSNLYSANFEYLVPANADPVVTQNIIVTGQGEV 817
           |||||||||||||:|||||||||||||||||||||||:||||||||||||||||||||||
Db    122  QFLLDADHNTFGSVIPAIGPLFTGTASSNLYSANFEYLIPANADPVVTQNIIVTGQGEV 181

Qy    818  VIPGGVYDYCITNPEPASGKMWIAGDGGNQOPARYDDFTFEAGKKYTFMRRAGMGDGTDM 877
           ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db    182  VIPGGVYDYCITNPEPASGKMWIAGDGDNQPARYDDFTFEAGKKYTFMRRAGMGDGTDM 241

Qy    878  EVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDV 937
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242  EVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDV 301

Qy    938  TVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASW 997
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302  TVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGT-TTLSESFENGIPASW 360

Qy    998  KTIDADGDGHGWKPGNAPGLAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDDLPNGGK 1057
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  KTIDADGDGHGWKPGNAPGLAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDDLANGGK 420

Qy   1058  LTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETTTAKGVRSPXAIRGRIQGTWR 1117
           ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db    421  LTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETTTAKGVRSPEAIRGRIQGTWR 480

Qy   1118  QKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTT 1177
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  QKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTT 540

Qy   1178  IDADGDGQWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNVLISKDVTGATKVKYY 1237
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  IDADGDGQWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYY 600

Qy   1238  YAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWI 1297
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  YAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWI 660

Qy   1298  ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLT 1357
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  ERTVDLPAGTKYVAFRHYNCSDLDYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLT 720

Qy   1358  ETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAIL 1417
           ||||||||||||||||||||||||||||||||||:|::|: |:|
Db    721  ETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTIINPTQFNPVKNLKAQ--PDGGDVVL 778

Qy   1418  KWNAPASKRAEVLNEDFE-NGIPASWKTIDADGDGNNW--TTTPPGG-----SSFAGHN 1469
           ||:|| :|:|| |:||||||:|:: :|::  ::|
Db    779  KWEAPSGKRGELLNEDFEGDAIPTGWTALDADGDGNNWDITLNEFTRGERHVLSPLRASN 838

Qy   1470  SAICVSSASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQD-ANYASEHYAVYASST 1528
           ||:|:|:||| |:|:|| ::  :|::||:|:
Db    839  VAISYSSLLQGQEYLPLTPNNFLITPKVE--GAKKITYKVGSPGLPQWSHDHYALCISKS 896

Qy   1529  GNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTD 1588
           |||| :| :||| :|| |:: ::|
Db    897  GTAAADF------EVIFEETMTYTQGGANLTREK------DLPAGTKYVAFRHYNCTD 942

Qy   1589  FFWINLDDVVIT-SGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPN 1647
           ||||||||||||||:|:|
Db    943  VLGIMIDDVVITGEGEGPSYTYTVYRDGTKIQEGLTETTYRDAGMSAQSHEYCVEVKYAA 1002
```

```
Qy   1648 GESAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVY 1707
               :|||||||||||||||||||||||||||||||||||||||||||||||
Db   1003 GVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVY 1062

Qy   1708 TAQGGHYAVMVVVDGKSYVEKLAVK 1732
               ||||:|||||||||||||||||:
Db   1063 TAQGGYYAVMVVVDGKSYVEKLAIK 1087
```

RESULT 9
AAW24786
ID   AAW24786 standard; protein; 1706 AA.
AC   AAW24786;
XX
DT   17-OCT-2003 (revised)
DT   25-NOV-1997 (first entry)
XX
DE   PrtR antigenic protein complex.
XX
KW   Periodontal disease; cell surface protein; thiol protease; endopeptidase;
KW   PrtR; PrtR45; PrtR44; PrtR15; PrtR17; PrtR27; haemagglutinin; adhesin;
KW   therapy; diagnosis; vaccine; antigen.
XX
OS   Porphyromonas gingivalis; strain W50.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..227
FT   /label= Pro-pro peptide
FT   Cleavage-site   227..228
FT   Protein   228..719
FT   /label= PrtR45
FT   /note= "45 kDa Arg-specific thiol protease"
FT   Cleavage-site   719..720
FT   Protein   720..1138
FT   /label= PrtR44
FT   /note= "44 kDa adhesin"
FT   Cleavage-site   1138..1139
FT   Protein   1139..1273
FT   /label= PrtR15
FT   /note= "15 kDa adhesin"
FT   Cleavage-site   1273..1274
FT   Protein   1274..1431
FT   /label= PrtR17
FT   /note= "17 kDa adhesin"
FT   Cleavage-site   1431..1432
FT   Protein   1432..1706
FT   /label= PrtR27
FT   /note= "27 kDa adhesin"
XX
PN   WO9716542-A1.
XX
PD   09-MAY-1997.
XX
PF   30-OCT-1996; 96WO-AU000673.
XX
PR   30-OCT-1995; 95AU-00006275.
XX
PA   (UYME) UNIV MELBOURNE.
PA   (VICT-) VICTORIAN DAIRY IND AUTHORITY.
XX
PI   Reynolds EC, Bhogal PS, Slakeski N;
XX
DR   WPI; 1997-272112/24.
DR   N-PSDB; AAT78850.
XX
PT   New antigenic protein complex from Porphyromonas gingivalis - comprising
PT   Arg- and Lys- specific thiol endo-peptidase(8), used in the detection,
PT   prevention and treatment of periodontal disease.
XX
PS   Example 1; Fig 8b; 68pp; English.
XX

CC   A PrtR-PrtK cell surface protein of Porphyromonas ginivalis (PG)
CC   comprises a 300 kDa complex composed a 45 kDa arginine-specific thiol
CC   protease and 44, 15, 17 and 27 kDa adhesins encoded by the prtR gene
CC   (AAT78850), and a 148 kDa lysine-specific thiol protease and 39, 15 and
CC   44 kDa adhesins (see AAW24787) encoded by the prtK gene (AAT78851). A
CC   claimed antigenic complex comprises at least one multimeric protein
CC   complex of PrtR and PrtK each containing at least one adhesin domain, the
CC   complex having a mol.wt. of over 200 kDa, and preferably comprises all 9
CC   proteins of the PrtR-PrtK complex (see also AAW24780-85). It can be used
CC   in a claimed composition to elicit an immune response directed against
CC   PG, and in a claimed method of reducing the prospect of PG infection
CC   and/or severity of disease. Antibodies directed against the complex are
CC   claimed for use in treating PG infection. Unlike whole PG cells or other
CC   previously prepared antigens based on fimbriae or the capsule, the PrtR-
CC   PrtK complex or component parts are safe and effective antigens. (Updated
CC   on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1706 AA;

Query Match   46.5%;
Best Local Similarity   50.9%;   Pred. No. 3.6e-274;
Matches 932; Conservative 204; Mismatches 456; Indels 240; Gaps 47;
                        Score 4272; DB 2; Length 1706;
```

```
Qy      6 LLIAASLLGVGLYAQSAKIKLDAPTRTTCNNSFKQFDASFSFNEVELTKVETKGGTFA   65
            : : |||| ||| | :|| | ||||||||||||||||||||:||||:|||
Db     10 IALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGI-   65

Qy     66 SVSIPGAFPT-------GEVGSPEVPAVRKLIAVPVGATPVVRVKS--FTEQVYSLNQYG  116
Db     66 -----GQVPTYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL----  116

Qy    117 SEKLMPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINP  176
Db    117 ---IAPSKGMIMRNEDPKKIPYVY-GKTYSQNKFFPGEIATLDDPFILRDVRGQVVNFAP  172

Qy    177 VQYDVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTD  233
Db    173 LQYNPVTKTLRIYTEITVAVSETSEQGKNILNKKGTFAG------FEDTYKRMF----MN  222

Qy    234 HGDLYNTPV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASI  287
Db    223 YEPGRYTPVEEKQNGMVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI    282

Qy    288 KAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYYSAVDGDYF  341
Db    283 QQFVKQEYEKEGND--------LTYVLLIGDHKDIPAKITPGIKSDQV--YGQIVGNDHY  332

Qy    342 PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK  401
Db    333 NEVFIGRFSCSESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEBGGPSADNGBSDIQ  392

Qy    402 Y-GMQYYYNQEHGYTDVYNYLKAPYTGCY----------SHLNTGVSFANYTAHGSETAW  450
Db    393 HENVIANLLTQYGYTKIIK------CYDPGVTPKNIIDAFNGGISLANYTHGHGSETAW   444

Qy    451 ADPLLTTSQLKALTNKDKYFLAIGNCCITAQPDYVQPCFGEVITRV---KEKGAYAYIG   506
Db    445 GTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFABALMRAQKDGKPTGTVAIIA  504

Qy    507 SSPNSYWGEDYWSVGANAVFGVQPTTPEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH  566
Db    505 STINQSW---------ASPMRG-------------QDEMNEI-------LCEKH       529

Qy    567 AGNI----GNITHIGAHYWEAYH--------VLGDGSVMPYRAMPKINTYLPASL      611
Db    530 PNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGPSLLVRTLVPTKMQVTRAPAQI  589

Qy    612 FQNQASYSIQAS-AGSVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN   670
Db    590 NLTDASVNVSCDYNGAIATISANGKMFGSAVVB-NGTATINLT-GLTNESTLTLTVVGYN  647

Qy    671 YLPVIKQIQV-GEPSPYQFVSNLTATTQGQKVTLKWEAPSAK---KAEGSREVKRIGDGL  726
```

```
RESULT 10
AAY67396
ID   AAY67396 standard; protein; 1704 AA.
XX
AC   AAY67396;
XX
DT   25-APR-2000  (first entry)
XX
DE   Arg-gingipain-2 amino acid sequence.
XX
KW   Arginine specific proteinase; Arg-gingipain; gingipain-2; haemagglutinin;
KW   immunogenic component; vaccine; inflammatory response; tissue damage;
KW   periodontal disease.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key             Location/Qualifiers
FT   Region          229..719
FT                   /note= "Amino acids 229-719 are specifically claimed"
FT   Region          720..1185
FT                   /note= "Amino acids 720-1185 are specifically claimed"
XX
PN   US6017532-A.
XX
PD   25-JAN-2000.
XX
PF   08-NOV-1994;  94US-00336308.
XX
PR   10-SEP-1993;  93US-00119361.
PR   24-JUN-1994;  94US-00265441.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Potempa JS,  Travis J;
XX
DR   WPI;  2000-136659/12.
DR   N-PSDB;  AAZ60181.
XX
PT   New Porphyromonas gingivalis arginine-specific protease preparation
PT   useful for preparing vaccines against periodontal disease and for
PT   screening for Arg-gingipain inhibitors.
XX
PS   Claim 1; Col 29-42; 55pp; English.
XX
CC   This sequence represents a Porphyromonas gingivalis arginine-specific
CC   proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
CC   Gingipain-2 consists or a 50kD protease component non-covalently
CC   associated with a 44kD haemagglutinin component. The proteinase is
CC   stimulated by glycine containing peptides and glycine analogues. It is
CC   inhibited by cysteine protease group specific inhibitors. The protease
CC   preparation can be used in immunogenic compositions and vaccines against
CC   inflammatory response and tissue damage caused by P. gingivalis in
CC   periodontal disease. It can also be used to screen for agents that
CC   modulate Arg-gingipain proteinase activity inhibitors.
XX
SQ   Sequence 1704 AA;

Query Match             46.5%;   Score 4270;  DB 3;  Length 1704;
Best Local Similarity   50.9%;   Pred. No. 4.9e-274;
Matches  932;  Conservative 204;  Mismatches 204;  Indels 456;  Gaps 238;
```

```
Db   648  KETVIKTINTNGEPNPYQPVSNLTATTGGQKVTLKWDAPSTKTNATTNTARSVDGIRELV  707

Qy   727  FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNIFGSVIPA-TGPLFTG-TA  783
Db   708  LLSVSDAPELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQYGQVIPSDTHTLWPNCSV  767

Qy   784  SSNLYSANFEYLVPANADPVVTQNIIVTGQGEVIPGGVYDYCITNPEPASGKMWIAGD    843
Db   768  PANLF-APFEYIVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQ  825

Qy   844  GGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLT  903
Db   826  G---PTKEDDYVFEAGKKYHFLMKKMGSDGTELTISEGGSDYTYTVYRDGTKIKEGLT    882

Qy   904  ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK  963
Db   883  ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK  942

Qy   964  WDAPNGTENPNPNPNPNPG---TTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY  1020
Db   943  WDAPNGTPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY    1002

Qy  1021  NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS  1080
Db  1003  NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS  1062

Qy  1081  TGNDASNFTNALLEETITAKGVRSPKAIRGRIQSTWRQKTVDLPAGTKYVAFRHFQSTDM  1140
Db  1063  TGNDASNFTNALLEETITAKGVRSPEAMRGRIQSTWRQKTVDLPAGTKYVAFRHFQSTDM  1122

Qy  1141  FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA  1200
Db  1123  FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA  1182

Qy  1201  HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA  1260
Db  1183  HGGTNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA  1242

Qy  1261  GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL  1320
Db  1243  GDFTVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL  1302

Qy  1321  NYILLDDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT   1380
Db  1303  NYILLDDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT   1362

Qy  1381  AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE----------   1428
Db  1363  AGVSPKKCVNVTVNSTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTEGSREVKRIGDGL  1420

Qy  1429  ---------VLNEDFENGIPASWKTIDADGDGNNWTTTPPPGGSSFAGHNSA         1471
Db  1421  FVTIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGGSVIPATGPLFTGTASS 1479

Qy  1472  ICVSSASHINFEG--PQNPD-----NYLVT--PELSLPCGGTLTFWVCAQDANYASEHY   1521
Db  1480  DLYSA----NFESLIPANADPVVTQNIIVTGQGEVVIPGG---VYDYCITNPEPASGKM   1532

Qy  1522  AVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-A  1580
Db  1533  WI-AGDGGNQPARYDDFTFE-----------------------AGKKYTFT          1559

Qy  1581  FRHFGCTDFFWINLLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDDLATGFYTG   1640
Db  1560  MRRAGMGDG----TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYC   1614

Qy  1641  VKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAA  1700
Db  1615  VEVKYTAGVSPKVCVCDYIPDGVADVTAQKPYTLITVGKTITVTCQGEAMIYDMNGRRLAA 1674

Qy  1701  GRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK  1732
Db  1675  GRNTVVYTAQGGYYAVMVVVDGKSYVEKLAIK  1706
```

```
Qy    6   LLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELIKVETKGGTFA  65
Db   10   IALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVTK--VQPRMDNLKFTEVQTPKGI--  65

Qy   66   SVSIPGAFPT------GEVGSPEVPAVRKLIAVPVGATPVPVRVKS--FTEQVYSLNQYG  116
Db   66   ----GQVPTYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL----  116

Qy  117   SEKLMPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINP  176
```

Db 117 ---IAPSKGMIRNEDPKKIPYVY-GKSYSQNKFFPGEIATLDDPFIIRDVRGQVVNFAP 172
Qy 177 VQYDVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVVTD 233
Db 173 LQYNPVTKTLRIYTEITVAVSETSEQGKNILNKKGTFAG-----FEDTYKRMF------MN 222
Qy 234 ------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASI 287
Db 223 YEPGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI 282
Qy 288 KAFIHKKY----NDGLAASAAPVFLALVGDTDVLSGE-KGKKTKKVTDLYYSAVDGDYF 341
Db 283 QQFVKQEYEKEGND--------LTYVLLVGDHKDIPAKITPGIKSDQV---YGQIVGNDHY 332
Qy 342 PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLBKVLLIAGADYSWNSQVGQPTIK 401
Db 333 NEVFIGRFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQ 392
Qy 402 Y-GMQYYYNQEHGYTDVYNYLKAPYTGCY---------SHLNTGVSPANYTAHGSETAW 450
Db 393 HENVIANLLTQYGYTKIIK------CYDPGVTPKNIIDAFNGGISLVNYTGHGSETAW 444
Qy 451 ADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV----KEKGAYAYIG 506
Db 445 GTSHFGTTHVKQLINSNQLPFIFDVACVNGDFLFSMPCFABALMRAQKDGKPTGTVAIIA 504
Qy 507 SSPNSYWGEDYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH 566
Db 505 STINQSW------ASPMRG--------QDEMNEI-----LCEKH 529
Qy 567 AGNI----GNITHIGAHYWEAYH--------VLGDGSVMPYRAMPKTNTYTLPASL 611
Db 530 PNNIKRTFGGVTMNGMFAMVEKYKKDGERKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQI 589
Qy 612 PQNQASYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN 670
Db 590 NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGYN 647
Qy 671 YLPVIKQIQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK---KAEGSREVKRIGDGL 726
Db 648 KETVIKTINTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELV 707
Qy 727 FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA 783
Db 708 LLSVSDAPELLRSGQAEIVLEAHDVWNDGSSYQILLDADHDQYGQVIPSDTHTLWPNCSV 767
Qy 784 SSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKWWIAGD 843
Db 768 PANLF-APFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQ 825
Qy 844 GGNQPARYDDFTFEAGKKYFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLT 903
Db 826 G---ETKEDDYVFEAGKKYHFLMKKWGSGDGTELTISEGGSDYTYTVYRDGTKIKEGLT 882
Qy 904 ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK 963
Db 883 ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK 942
Qy 964 WDAPNGTPNPNPNPNPG-TTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNS 1022
Db 943 WDAPNGTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNS 1002
Qy 1023 NGCVYSESFGLGGIGVLTPDNYLITPALDLNGGKLTFWVCAQDANYASEHYAVVASSTG 1082
Db 1003 NGCVYSESFGLGGIGVLTPDNYLITPALDLNGGKLTFWVCAQDANYASEHYAVVASSTG 1062
Qy 1083 NDASNFTNALLEETIAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY 1142
Db 1063 NDASNFTNALLEETIAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY 1122
Qy 1143 IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDQGWLCLSSGQLDWLTAHG 1202
Db 1123 IDLDEVEIKANGKRADFTETFESSTHGEAPARWTTIDADGDQGWLCLSSGQLDWLTAHG 1182

Qy 1203 GSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGD 1262
Db 1183 GTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGD 1242
Qy 1263 FTVVFEETPNGINKGGARFGLSTEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNY 1322
Db 1243 FTVVFEETPNGINKGGARFGLSTEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNY 1302
Qy 1323 ILLDDIQFTMGGSPTPDTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG 1382
Db 1303 ILLDDIQFTMGGSPTPDTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG 1362
Qy 1383 VSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE------------ 1428
Db 1363 VSPKBCVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTBGSREVKRIGDGLFV 1420
Qy 1429 ------------VLNEDFENGIPASWKTIDADGDGNNWTTPPGGSSFAGHNSAIC 1473
Db 1421 TIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSNL 1479
Qy 1474 VSSASHINFEG--PQNPD-----NYLVT-PELSLPGGGTLTFWVCAQDANYASEHYAV 1523
Db 1480 YSA----NFEYLIPANADPVVTTQNIIVTGQGEVVIPGG---VIDYCITNPEPASGKMWI 1532
Qy 1524 YASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAAQGTWYQKTVQLPAGTKVV-AFR 1582
Db 1533 -AGDGNQPARYDDFTFE-------------------------AGKKYTFTMR 1559
Qy 1583 HFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLLATGFYTYGVK 1642
Db 1560 RAGMGDG---------------TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVE 1614
Qy 1643 VVYPNGESAIETATLNITSLADVTAQKPYTLTVGKTITVTCQGEAMIYDMNGRRLAAGR 1702
Db 1615 VKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGR 1674
Qy 1703 NTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
Db 1675 NTVVYTAQGGYYAVMVVVDGKSYVEKLAVK 1704

RESULT 11
AAU08938
ID   AAU08938   standard; protein; 1704 AA.
XX
AC   AAU08938;
XX
DT   18-DEC-2001 (first entry)
XX
DE   P. gingivalis high molecular weight Arg-gingipain-2.
XX
KW   Periodontitis; antiinflammatory; Arg-gingipain-2; AG-2; immunogen.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..227
FT   /label= Prepro_peptide
FT   Region   599..619
FT   /note= "Region of homology with cysteine proteases"
FT   Region   670..674
FT   /label= Proteolytic_component
FT   Cleavage-site 719
FT   Protein   720..1091
FT   /label= HGP_44kDa
FT   Cleavage-site 1091
FT   Protein   1092..1429
FT   /note= "Haemagglutinin protein component"
FT   Cleavage-site 1429
FT   Protein   1430..1704
FT   /note= "Haemagglutinin protein component"

```
FT        /label= HGP 27kDa
FT        /note= "Haemagglutinin protein component"
XX
PN    US6274718-B1.
XX
PD    14-AUG-2001.
XX
PF    25-JAN-2000; 2000US-00490931.
XX
PR    10-SEP-1993; 93US-00119361.
PR    24-JUN-1994; 94US-00265441.
PR    09-SEP-1994; 94WO-US010283.
PR    08-NOV-1994; 94US-00336308.
XX
PA    (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI    Travis J, Potempa JS, Barr PJ, Pavloff N;
XX
DR    WPI; 2001-588904/66.
DR    N-PSDB; AAS15242.
XX
PT    New recombinant DNA molecule which encodes high molecular weight (mature)
PT    Arg-gingipain protein, useful for immunization against inflammation and
PT    tissue damage, comprises enzymatically active protease component and
PT    hemagglutinin component.
XX
PS    Claim 1; Col 29-41; 56pp; English.
XX
CC    The invention relates to a recombinant DNA molecule encoding high
CC    molecular weight (mature) Arg-gingipain (AG) protein, which has an
CC    enzymatically active protease component (AG-2) and a haemagglutinin
CC    component, from P. gingivalis. The nucleic acid is useful for producing
CC    mature Arg-gingipain protein. Immunogenic compositions comprising Arg-
CC    gingipain are useful for immunising animals including humans against
CC    inflammatory response and tissue damage caused by an archaebacterium
CC    Porphyromonas gingivalis, which causes progressive periodontitis. Arg-
CC    gingipain is also useful for identifying agents that modulate Arg-
CC    gingipain proteinase activity, whether by acting on the proteinase itself
CC    or preventing the interaction of the proteinase with the protein in the
CC    gingival area, such as complement factors C3 or C5. The present sequence
CC    is Arg-gingipain-2
XX
SQ    Sequence 1704 AA;

Query Match        46.5%;   Score 4270;   DB 4;   Length 1704;
Best Local Similarity  50.9%;  Pred. No. 4.9e-274;
Matches  932;  Conservative 204;  Mismatches 456;  Indels 238;  Gaps  47;
```

```
Qy      6 LLIAASLLGVGLYAQSAKIKLDAPFTRTTCTNNSFKQFDASFSFNEVELTKVETKGGTFA   65
            :  :||||||  ::  : : ::    :  |     :: :  :|:  :||:|
Db     10 IALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGI--   65

Qy     66 SVSIPGAFPT--------GEVGSPEVPAVRKLIAVPVGATPVVRVKS--FTEQVYSLNQYG  116
                             |  :||| |  :|||: :|   :  : |  ||: :
Db     66 ----GQVPFTYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL----  116

Qy    117 SEKLMPHQPSMSKSDDPEKVPFVVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINP  176
            |:  ::|:|:|  ||:|  :  : :  : ::   |
Db    117 ---IAPSKGMIMRNEDPKKIPYVY-GKSYSQNKFFPGEIATLDDPFILRDVRGQVVNFAP  172

Qy    177 VQYDVVANQLKVRNNIEI--EVSFQGADEVATQRLYDASFSPYPETAYKQLFNRDVYTD  233
            ||: | :||  |        ::  : :   ::   :  ||  ::|
Db    173 LQYNPVTKTLRIYTBITVAVSETSEQGKNILNKKGTFAG-----FEDTYKRMF------MN  222

Qy    234 HGDLYNTPV-----RMLVVAGAKFKEALKPWLTWKAQGFYLDVHYTDRAEVGTTNASI  287
              :                ||| : ||  :|:||
Db    223 YEPGRYTPVPVEEKQNGRMIVTVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI  282

Qy    288 KAFIHKKY-----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYYSAVDGDYF  341
            :  ::|| |     ::  ||| :|  :|||
Db    283 QQFVKQEYEKEGND-------LTYVLVGDHKDIPAKITPGIKSDQV---YGQIVGNDHY  332

Qy    342 PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK  401
```

```
Db    333 NEVFIGRFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQ  392
            |::: |     ||: :  :|:  ||:
Qy    402 Y-GMQYYNQEHGYTDVYNIKAPYTGCY----------SHLNTGVSFANYTAHGSETAW  450
            : |:|         ||:|:|:
Db    393 HENVIANLLTQYGYTKIIK------CYDPGVTPKNIIDAFNGGISLVNYTGHGSETAW  444

Qy    451 ADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV----KEKGAYAYIG  506
            |:||| :  :  |::  ::| ||::  :||   |  : ||::
Db    445 GTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIA  504

Qy    507 SSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH  566
            |: |||
Db    505 STINQSW-----------------ASPMRG------------------QDEMNEI----LCEKH  529

Qy    567 AGNI----GNITHIGAHYYWEAYH------------VLGDGSVMPYRAMPKTNTYTLPASL  611
                    :  :                |::
Db    530 PNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQI  589

Qy    612 PQNQASYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN  670
            |::|:  : :   ||: :  :|| || | :| :
Db    590 NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLITLTVVGYN  647

Qy    671 YLPVLKQIQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK--KAEGSREVKRIGDGL  726
            |||  : ||: |||  |  |||:|||:|||:||:|   ||:|
Db    648 KETVIKTINTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELV  707

Qy    727 FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA  783
            |   :     |: ::  : :::   |      |::| :|:  ||  |||
Db    708 LLSVSDAPELLRSGQAEIVLEAHDVWNDGSSYQILLDADHDYGQVIPSDTHTLWPNCSV  767

Qy    784 SSNIYSANFEYLVPANADPVVTTQNIIVTQGEVVIPGGVYDYCIINPEPASGKMWIAGD  843
            |:    :||     :|   |    ||:    ||::||| : ||  || :  ||  |||
Db    768 PANLF-APFEYTVPENADPSCCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQ  825

Qy    844 GGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYVYRDGTKIKEGLT  903
            |   |     |  ||::|:|:|           |:::| :   :||:|||:|:|:|||
Db    826 G---PTKEDDYVFEAGKKYHFLMKKMGSGDGTELTISEGGSGDYTYTVYRDGTKIKEGLT  882

Qy    904 ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK  963
            |||||||||  ||||||||||||||||||| ||||||||||||||||||||  |||||||
Db    883 ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK  942

Qy    964 WDAPNGTPNPNPNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNS  1022
            ||||||||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db    943 WDAPNGTPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNS  1002

Qy   1023 NGCVYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG  1082
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1003 NGCVYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG  1062

Qy   1083 NDASNFTNALLEETTTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY  1142
            ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   1063 NDASNFTNALLEETTTAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY  1122

Qy   1143 IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWILCLSSGQLDWLTAHG  1202
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1123 IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWILCLSSGQLDWLTAHG  1182

Qy   1203 GSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGD  1262
            |: ||:||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1183 GTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGD  1242

Qy   1263 FTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNY  1322
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1243 FTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNY  1302

Qy   1323 ILLDDIQFTMGGSPTPTDYTYVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAG  1382
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1303 ILLDDIQFTMGGSPTPTDYTYVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAG  1362

Qy   1383 VSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE-----------  1428
            ||::|   |:|  :|| |  : ::   |    :  : : |
```

```
XX
PS   Disclosure; Page 70-77; 89pp; English.
XX
CC   A low mol.wt. arginine-specific gingipain (AG-1) and high mol.wt. AG (AG-
CC   2) were isolated from P. gingivalis strains H66 (ATCC 33277) and W50
CC   (ATCC 53973). The sequences of the proteins were used to design PCR
CC   primers and probes to isolate AG DNA. Lambda DASH and lambda ZAP
CC   libraries were screened with a probe based on amino acids 11-22 of the AG
CC   protein to obtain DNA encoding AG-1 (AAQ83484) and AG-2 (AAQ83489). AG-2
CC   is a prepolyprotein incorporating AG-1. (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 1704 AA;

Query Match            46.5%;   Score 4264;   DB 2;   Length 1704;
Best Local Similarity  50.9%;   Pred. No. 1.2e-273;
Matches 931;   Conservative 204;   Mismatches 457;   Indels 238;   Gaps 47;

Qy     6   LLIRAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETKGGTFA   65
           : :.:||  :  : .:. .:   ::. ::   :.  :.  .::
Db    10   IALCSSLLGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGI-     65

Qy    66   SVSIPGAFPT------GEVGSPEVPAVRKLIAVPGATPVVRVKS--FTEQVISLNQYG    116
Db    66   -----GQVPTYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL----  116

Qy   117   SEKLMPHQPSMSKSDDPEKVPFVYNAAAYARKGPVGQELTQVEMLGTMRGVRIAALTINP  176
Db   117   ---IAPSKGMIRNEDPKKLIPYVV-GKSYSQNKFFPGEIATLDDPFILRDVRGQVVNFAP  172

Qy   177   VQYDVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTD  233
Db   173   LQYNPVTKTLRIYTEITVAVSETSEQGKNILNKKGTPAG----PEDTYKRMF----MN   222

Qy   234   HGDLYNTPV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASI  287
Db   223   YEPGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI  282

Qy   288   KAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYSAVDGDYF  341
Db   283   QQFVKQEYEKEGND----------LTYLLVGDHKDIPAKITPGIKSDQV---YGQIVGNDHY  332

Qy   342   PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK  401
Db   333   NEVFIGRFSCSCSKEDIKTQIDRTIHYERNITTEDKWLGQALCIASABGGPSADNGESDIQ  392

Qy   402   Y-GMQYYNQEHGYTDVYNYLKAPYTGCY-------SHLNTGVSFANYTAHGSETAW      450
Db   393   HENVIANLLTQYGYTKIIK--------CYDPGVTPKNIIDAFNGGISLVNYTGHGSETAW  444

Qy   451   ADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV----KEKGAYAYIG  506
Db   445   GTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIA  504

Qy   507   SSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH  566
Db   505   STINQSW--------ASPMRG-------------QDEMNEI--------LCEKH      529

Qy   567   AGNI-----GNITHIGAHYYKEAYH----------VLGDGSVMPYRAMPKTNTYTLPASL  611
Db   530   PNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFADPSLLVRTLVPTKMQVTAPAQI  589

Qy   612   PQNQASYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN  670
Db   590   NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGYN  647

Qy   671   YLPVIKQIQV-GEPSFYQPVSNLTATTQGQKVTLKWEAPSAK--KAEGSREVKRIGDGL   726
Db   648   KETVIKTINTNGEPNYPQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELV  707

Qy   727   FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA  783
Db   708   LLSVSDAPELLRSGQAEIVLEAHDVWNDGSSYQILLDADHDQYGQVIPSDHTLWPNCSV   767

Db  1363   VSPKECVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTEGSREVKRIGDGLFV  1420
Qy  1429   -----------------------VLNEDFENGIPASWKTIDADGDGNNWTTPPGGSSFAGHNSAIC  1473
Db  1421   TIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSNL  1479
Qy  1474   VSSASHINFEG--PQNPD------NYLVT-PELSLPGGGTLITPWVCAQDANYASEHYAV  1523
Db  1480   YSA----NFEYLIPANADPVVTTQNIIVTGQGEVIPGG---VYDYCIINPEPASGKMWI   1532
Qy  1524   YASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAGGTWYQKTVQLPAGTKYV-APR  1582
Db  1533   -AGDGGNQPARYDDFTFE-------------------------AGKKYTFTMR       1559
Qy  1583   HFGCTDFFWINLDDVVITSGNAPSYTYIYRNNTQIASGVTETTYRDPDLATGPFYTYGVK  1642
Db  1560   RAGMGDG-----TDMEVEDDSPASYTYTVVRDGVTKIKEGLTETTYRDAGMSAQSHEYCVE  1614
Qy  1643   VVYPNGESAIETATLNITSLADYTAQKPYTLTVGKTITVTCQGEAMIYDMNGRRLAAGR  1702
Db  1615   VKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGR  1674
Qy  1703   NTVVYTAQGGHYAVMVVVDGKSYYEKLAVK  1732
Db  1675   NTVVYTAQGGYYAVMVVVDGKSYYEKLAVK  1704

RESULT 12
AAR70188
ID   AAR70188 standard; protein; 1704 AA.
XX
AC   AAR70188;
XX
DT   25-MAR-2003  (revised)
DT   21-SEP-1995  (first entry)
XX
DE   Arg-gingipain-2 prepolyprotein.
XX
KW   Arg-gingipain-2; gingivalis; periodontal disease; vaccine;
KW   arginine-specific protease.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key            Location/Qualifiers
FT   Protein        228..719
FT                  /label= Protease
FT   Region         720..1091
FT                  /note= "corresponds to Arg-gingipain-1"
FT   Region         1092..1429
FT                  /label= Hemagglutinin
FT   Region         1430..1704
FT                  /label= Hemagglutinin
XX
PN   WO9507286-A1.
XX
PD   16-MAR-1995.
XX
PF   09-SEP-1994; 94WO-US010283.
XX
PR   10-SEP-1993; 93US-00119361.
PR   21-OCT-1993; 93US-00141324.
PR   24-JUN-1994; 94US-00265441.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J, Potempa J, Barr PJ, Pavloff N;
XX
DR   WPI; 1995-123373/16.
DR   N-PSDB; AAQ83489.
XX
PT   DNA encoding Arg-gingipain proteins - used to develop prods. for
PT   detection, treatment and prevention of periodontal disease.
```

XX
AC   AAW34843;
XX
DT   03-JUN-1998 (first entry)
XX
DE   Arg-gingipain high molecular weight prepolyprotein sequence.
XX
KW   Arg-specific gingipain protease; gingivalis; periodontal disease;
KW   vaccine; infection.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key   Location/Qualifiers
FT   Protein   1..227
FT   /note= "precursor protein"
XX
PN   WO9734629-A1.
XX
PD   25-SEP-1997.
XX
PF   21-MAR-1997;   97WO-US004635.
XX
PR   22-MAR-1996;   96US-0013945P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
PA   (MORE-) MOREHOUSE SCHOOL MEDICINE.
XX
PI   Potempa J, Travis J, Genco C;
XX
DR   WPI; 1997-479993/44.
DR   N-PSDB; AAT93872.
XX
PT   Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
PT   useful for protecting animals and humans from gingivalis and periodontal
PT   diseases.
XX
PS   Disclosure; Page 68-73; 95pp; English.
XX
CC   The present sequence represents an arginine-specific protease of
CC   Porphyromonas gingivalis. The following peptides, derived from Arg- and
CC   Lys-specific high molecular weight proteases, offer protection against
CC   infection: YYTTVYRDGK IKEGLTATTE DDGVATGNHE YCVEKYTAGS VSPKVC (I);
CC   YTPVEEKQNG RMVIVAKKY (II); QLPFIFDVAC VNGDFLFSMP CFAEALMRAQ (III);
CC   GEPNPYQPVS NLTATTQGQK VTLKWDAPSTK (IV); GNHEYCVEVK YTAGVSPKVC KDVTV (V);
CC   RMFMNYEPGR YTPVEEKQNG (VI); TFAGPEDTYK RMFMNYEPGR (VII); DYTYTVYRDG
CC   TKIKEGLTAT TFEEDGVATG NMEYCVCVKY TAGVSPKVC (VIII); YYTTVYRDGT KIKEGLTATTF
CC   EEDG (IX); RDGTKIKEGL TATTFEEDGV ATGN (X); KIKEGLTATT FEEDGVATGN HEY (XI)
CC   ; KWDAPNGTPN PNPN PNPNPN PGTTTLSE (XII); and YTPVEEKENG RMVIVAKKY
CC   (XIII). They are used in vaccines to protect animals, including humans,
CC   from gingivitis and/or periodontal diseases
XX
SQ   Sequence 1704 AA;

Query Match   46.4%; Score 4263; DB 2; Length 1704;
Best Local Similarity   50.9%; Pred. No. 1.4e-273;
Matches 931; Conservative 204; Mismatches 457; Indels 238; Gaps 47;

Qy   6 LLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSNEVELTKVETKGGTFA 65
            .:|:||||| |||| : :|| ::|:: :|| |::   :|| :::||::|||
Db   10 IALCSSLLGSMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGI- 65

Qy   66 SVSIPGAFPT--------GEVGSPEVPAVRKLIAVPVGATPVPVRVKS--FTEQVSLNQYG 116
           :||:||: | ||:: |||||| | :::|| |: | || :|||:|:||
Db   66 ----GQVPYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL---- 116

Qy   117 SEKLMPHQPSMSKSDDPEKYPFVYNAAAYARKGVGQELTQVEMLGTMRGVRIAALTINP 176
           | :||: :| ::| :|::| | : :: :   :| :: ||
Db   117 ---IAPSKGMIMRNEDPKKIPYV--GKSYSQNKFFPGEIATLDDPFILRDVRGQVVNFAP 172

Qy   177 VQYDVVANQLKVRNNIEI--EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTD 233
           :||:|| || :|:|::   :|::| | |:| |::::|:   :|
Db   173 LQYNPVTKILRIYTEITVAVSETSEQGKNILNKKGTFAG----FEDTYKRMF-----MN 222

RESULT 13
AAW34843
ID   AAW34843 standard; protein; 1704 AA.

Qy	784	SSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVDYCIINPEPASGKMWIAGD	843
Db	768	PANLF-APFEYIVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIIAAPQ-ANAKIWTAGQ	825
Qy	844	GGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTTVYRDGTKIKEGLT	903
Db	826	G---PTKEDDYVFEAGKKYHFLMKKMGSGDGTELTISEGGGSDYTTVYRDGTKIKEGLT	882
Qy	904	ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK	963
Db	883	ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK	942
Qy	964	WDAPNGTPNPNPNPNPNPG-TTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNS	1022
Db	943	WDAPNGTPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNS	1002
Qy	1023	NGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG	1082
Db	1003	NGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG	1062
Qy	1083	NDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY	1142
Db	1063	NDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY	1122
Qy	1143	IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHG	1202
Db	1123	IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHG	1182
Qy	1203	GSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGD	1262
Db	1183	GTNVVASFSWNGMALNPDNYLISKDVTGATKVKCYYAVNDGFPGDHYAVMISKTGTNAGD	1242
Qy	1263	FTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNY	1322
Db	1243	FTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNY	1302
Qy	1323	ILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAG	1382
Db	1303	ILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAG	1362
Qy	1383	VSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE---------------	1428
Db	1363	VSPKECVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTEGSREVKRIGDGLFV	1420
Qy	1429	----------VLNEDFENGIPASWKTIDADGDGNNWTTTPPPGGSSFAGHNSAIC	1473
Db	1421	TIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSNL	1479
Qy	1474	VSSASHINFEG--PQNPD------NYLVT---PELSLPGGGTLTFWVCAQDANYASEHYAV	1523
Db	1480	YSA-----NFEYLIPANADPVVTTQNIIVTGQGEVVIPGG---VDYCIINPEPASGKMWI	1532
Qy	1524	YASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-APR	1582
Db	1533	-AGDGGNQPARYDDFTFE------------------------AGKKYTFTMR	1559
Qy	1583	HFGCTDFFWINLDDVVITSGNAPSYTYTYRNNTQIASGVTETTYRDPDLATGFYTGVK	1642
Db	1560	RAGMGDG------TDMEVEDDSPASYTTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVE	1614
Qy	1643	VVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGR	1702
Db	1615	VKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGR	1674
Qy	1703	NTVVYTAQGGHYAVMVVVDGKSYVEKLAVK	1732
Db	1675	NTVVYTAQGGYYAVMVVVDGKSYVEKLAVK	1704

```
Db  1243 FTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNY 1302
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1323 ILLDDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG  1382
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1303 ILLDDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG  1362
Qy  1383 VSPKKCVDVTVNSTQFNPVQNLTABQAPNSMDAILKWNAPASKRAE------------ 1428
         |||||:|||||:|:||:||||:|||||| :|::|||: :
Db  1363 VSPKECVNVTINPTQFNPVQNLKAQ--PDGGDVVLKWEAPSAKKTEGSREVKRIGDGLFV 1420
Qy  1429 ---------------VLNEDFENGIPASWKTIDADGDGNNWTTPPPGGSSFAGHNSAIC 1473
                        :|:
Db  1421 TIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSNL 1479
Qy  1474 VSSASHINFEG--PQNPD------NYLVT--PELSLPGGGTLTFWVCAQDANYASEHYAV 1523
         |:  ||||||| :|:          ||||   |||||||||||||||||
Db  1480 YSA----NFEYLIPANADPVVTTQNIIVTGQGEVVIPGG---VDYCITNPEPASGKMWI 1532
Qy  1524 YASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-AFR 1582
         ||| ||
Db  1533 -AGDGGNQPARYDDFTFE---------------------------AGKKYTFTMR    1559
Qy  1583 HFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVK 1642
         |:|:|||:| ||:|||||||:||||||||| |||:||:|| ||||||| :||||:|||||
Db  1560 RAGMGDG---TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVE   1614
Qy  1643 VVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTIIVTCQGEAMIYDMNGRRLAAGR 1702
         |:||| :|||||||||||||:|||||||||||||||||||||:|||||||||||| ||||
Db  1615 VKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRGLAAGR 1674
Qy  1703 NTVVYTAQGGHYAVMVVVDGKSYVEKIAVK 1732
         |||||||||| |||||||||||||||:||||
Db  1675 NTVVYTAQGGYYAVMVVVDGKSYVEKIAVK 1704
```

```
RESULT 14
AAW69495
ID  AAW69495 standard; protein; 1687 AA.
XX
AC  AAW69495;
XX
DT  22-DEC-1998  (first entry)
XX
DE  Haemagglutinin protein hagE.
XX
KW  Haemagglutinin protein; periodontal disease; vaccine; hagE.
XX
OS  Porphyromonas gingivalis.
XX
PN  US5824791-A.
XX
PD  20-OCT-1998.
XX
PF  11-DEC-1995;            95US-00570311.
XX
PR  08-SEP-1988;            88US-00241640.
PR  25-JAN-1991;            91US-00647119.
PR  09-DEC-1994;            94US-00353485.
XX
PA  (UYFL ) UNIV FLORIDA.
PA  (UABR-) UAB RES FOUND.
XX
PI  Patti JM,  Han N,  Lantz M,  Tumwasorn S,  Progulske-Fox A,  Lepine G;
XX
DR  WPI; 1998-582627/49.
DR  N-PSDB; AAV58881.
XX
PT  Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT  protease poly:peptide(s).
XX
PS  Claim 1; Col 167-182; 101pp; English.
XX
```

```
Qy   234 HGDLYNTPV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASI   287
         ::|:  |:::     : ||:| |||||::|::||||| |    |:   |  :::
Db   223 YEPGRYTPVEEKQNGRMVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI   282
Qy   288 KAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE-KGKKTKKVTDLYYSAVDGDYF   341
         :|   |      :: |||||||:||:| :||||::||  |    ::|:|||:
Db   283 QQFVKQEYEKEGND------LTYVLVGDHKDIPAKTPGIKSDQV---YGQIVGNDHY    332
Qy   342 PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK  401
         |   :|||:::|:|||::::|: ||::: :|| :||:|:|||:|||:|||||:||||::|
Db   333 NEVFIGRFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGPSADNGESDIQ   392
Qy   402 Y-GMQYYNQEHGVTDVYNYLKAPYTGCY-----------SHLNTGVSFANYTAHGSETAW  450
         |  |: :  |||:| ::|:|||||||                |::||:|||||||||||
Db   393 HENVIANLLTQYGYTKIIK--------CVDPGVTPKNIIDAFNGGISLVNYTGHGSETAW  444
Qy   451 ADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV----KEKGAYAYIG  506
         :|||| ||:|:|:|:|:|:||:|||||:||||||||||||||||||     |   |||||
Db   445 GTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGVTAIIA  504
Qy   507 SSPNSYWGEDYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH   566
         :::|:       |        ||||:
Db   505 STINQSW-------ASPMRG-----------QDEMNEI------LCEKH           529
Qy   567 AGNI----GNITHIGAHYWEAYH-----------VLGDGSVMPYRAMPKTNTYTLPASL   611
         |      |:  |  |:||::          |:|||||:|||||||||| ||||||||
Db   530 PNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQI  589
Qy   612 PQNQASYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN  670
         : ::::::::|:  ::||:|::|| :|  |||:|  :|:|::||::||:|||||:|:|||
Db   590 NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGYN  647
Qy   671 YLPVIKQIQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK---KAEGSREVKRIGDGL 726
         :|:|:|::      ::|||||||||||||||||||||||    |||||||||||||||
Db   648 KETVIKTINTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELV  707
Qy   727 FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA  783
         ::||||| | :||:|||||:||:|:::|||:|||| ||||||:|||||  |||:||| ||
Db   708 LLSVSDAPELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQYGGQVIPSDTHTLWPNCSV  767
Qy   784 SSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVDYCITNPEPASGKMWIAGD  843
         ||||||:||| ||||:|::|||||||||||||||||| ||||||||||||||||||||
Db   768 PANLF-APFEYTVPENADPSCSSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQ  825
Qy   844 GGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLT  903
         |   ||||:: :: | ::||::
Db   826 G---PIKEDDYVFEAGKKYHFLMKKMGSGDGTELTISEGGSDYTYTVYRDGTKIKEGLT   882
Qy   904 ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK  963
         |||||||||:|||||||||||||||||||||||||||||||||||||||||:|||||||||
Db   883 ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK  942
Qy   964 WDAPNGTPNPNPNPNPG--TTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNS   1022
         |||||||||||||||     ||||||||||||||||||||||:|||||||||||||||
Db   943 WDAPNGTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNS   1002
Qy  1023 NGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG  1082
         ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db  1003 NGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG  1062
Qy  1083 NDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY  1142
         ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db  1063 NDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY  1122
Qy  1143 IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHG  1202
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1123 IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHG  1182
Qy  1203 GSNVVSSFSWNGMALNPDNYLISKDVTGATKYKCYYAVNDGFPGDHYAVMISKTGTNAGD  1262
         |:|:||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1183 GTNVVASFSWNGMALNPDNYLISKDVTGATKYKCYYAVNDGFPGDHYAVMISKTGTNAGD  1242
Qy  1263 FTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNY  1322
```

```
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagE haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1687 AA;

Query Match       46.2%;   Score 4245;   DB 2;   Length 1687;
Best Local Similarity  51.6%;   Pred. No. 2.2e-272;
Matches 920;   Conservative 197;   Mismatches 443;   Indels 224;   Gaps 44;

Qy    47  FSFNEVELTKVEKGGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKS--   104
           :| ||::|::: : ||::||: :|: : ||:|::|  :||:::::::|||| :|||
Db    30  FRMDNLKFTEVQTPKGMAQVPTYTEGVNLSEKGMPTLPIISRSLAVSDTREMKVEVVSSK   89

Qy   105  FTEQVYSLNQYGYGSEKLMPHQPSMSKSDDPEKVPFYVNAAAYARKGFVGQELTQVEMLGTM  164
           :|  ::  :  :| ::  : :::|:  || ||:| |  :||
Db    90  FIEKKNVL------IAPSKGMIMRNEBDPKKIPYVY-GKSYSQNKFFPGEIATLDDPFL   141

Qy   165  RGVRIAALTINPVQYDVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETA   221
           | ::: :|||:  :::  ::::   ::::     :: :||:::   :::
Db   142  RDVRGQVVNFAPLQYNPVTKTLRIYTEITVAVSETSEQGKNILNKKGTFAG------FEDT  196

Qy   222  YKQLFNRDVYTDHGDLYNTPV-----RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYT   275
           || :   :  | :| |||  ::::    ::|||:||  :::  ::||:
Db   197  YKRMF-----MNYEPGRYTPVEEKQNGRMFVIVAKKYEGDIKDFVDWKNQRGLRTEVKVA   251

Qy   276  DEAEVGTTNASIKAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVT   329
           :   ::  | ::: ::  :|    ||| :|:  ::::|:|||: :     ||:|:|:|
Db   252  EDIASPVTANAIQQFVKQEYEKEGND-----LTYVLLVGDHKDIPAKITPGIKSDQV-   303

Qy   330  DLYYSAVDGDYFPEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADY   389
           ::     ||::|  ::  |:  :|:||| :|:||: : ||:  ::|||::::|:|:||
Db   304  --YGQIVGNDHYNEVFIGRFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEG   361

Qy   390  SWNSQVGQPTIKY-GMQYYYNQEHGYTDVYNYLKAPYTGCY---------SHLNTGVSF   438
           :: :|  |:  || |:|| |::| ||:  :  :|:::           |:::||::|
Db   362  GPSADNGESDIQHENVIANLLTQYGYTKIIK------CYDPGVTPKNIIDAFNGGISL   413

Qy   439  ANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQPDYVQPCFGEVITRV--   496
           .||| |||||||. ::| ||:|:||||.::::: : | |:: :|:|  |:|  :|:
Db   414  VNNTGHGSETAWGTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFABALMRAQK   473

Qy   497  --KEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVN   554
             :|::     :  ||:    :  ::||:::  |::
Db   474  DGKPTGTVAIIASTINQSW---------ASPMRG------QDEMNEI-   505

Qy   555  SIMWAGNLAATHAGNI----GNITHIGAHYWEAYH-------VLJGDGSVMPYRAM   599
           :: :||: || | :|     ||:| : |:|:
Db   506  ----LCEKHPNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLV   558

Qy   600  PKTNTYTLPASLPQNQASYSIQAS-AGSVYAISKDGVLYGTGVANASGVATVSMTKQITE   658
           | : : : ||:| :| ||:: :|:  || :| ::| :  :| ||: :|:||::: :::
Db   559  PTKMQVTAPAQINLTDASVNNVSCDYNGAIATISANGKMFGSAAVE-NGTATINLT-GLTN  616

Qy   659  NGNYDVVITRSNYLPVIKQIQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK--KAE   714
           .| ::::|| :: :::|::: : ||:.|||||||||||||||||||||:|||:|   ||
Db   617  ESTLTLTVVGVNKEVIKTINTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKINATTN   676

Qy   715  GSREVKRIGDGLFVTIEPAND-VRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIP   773
           ::| ::|||||||||||||:| ||||||:|:||||:|:|||||| :|||||: :|||||
Db   677  TARSVDGIRELVLLSVSDAPELLRSGQAEIVLERAHDVWNDSGYQILLDADHDQYGQVIP   736

Qy   774  A-TGPLFTG-TASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNP   831
           :  | : ||  :||||: | ||| ||||||| :::  :: |:  |:: :| |||  |:||
Db   737  SDTHTLWPNCSVPANLF-APFEYTVPENADPSCCSPTNMIMDGTASVNIPAGTYDFAIAAP  795

Qy   832  EPASGKMWIAGDGGNQOPARYDDFTFBAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTV   891
           : ::||:|||| | :  ||||||||| ||||| | |:: |:||| | : :  :|||||||
Db   796  Q-ANAKIWIAGQG---PTKEDDYVFEAGKKYHFLMKKMGSGDGTELTISEGGSGSDYTYTV   851

Qy   892  YRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNL   951
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   852  YRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNL   911

Qy   952  TGSSVGQKVTLKWDAPNGTPNPNPNPNPG---TTLSESFENGIPASWKTIDADGDGHG   1008
           ||| |||||||||||||||||||||||:|   |||||||||||||||||||||||||
Db   912  TGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHG   971

Qy  1009  WKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDAN  1068
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   972  WKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDAN  1031

Qy  1069  YASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTK  1128
           |||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db  1032  YASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTK  1091

Qy  1129  YVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL  1188
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1092  YVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL  1151

Qy  1189  CLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDH  1248
           |||||||||||||||:|||:|||||||||||||||||||||||||||||||||||||||
Db  1152  CLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDH  1211

Qy  1249  YAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTK  1308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1212  YAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTK  1271

Qy  1309  YVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYTYTYVYRDGTKIKEGLTETTFEEDGVAT  1368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1272  YVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYTYTYVYRDGTKIKEGLTETTFEEDGVAT  1331

Qy  1369  GNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE  1428
           |||||||||||||||||:||:||| |:|||||:||: ::  | : :::|:|:||:||::|
Db  1332  GNHEYCVEVKYTAGVSPKECVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTE  1389

Qy  1429  --------------------VLNEDFENGIPASWKTIDADGDGNNWTTTPP  1459
                               ::.||:|||:   |: :  ||::: :
Db  1390  GSREVKRIGDGLFVTIEPANDVRANEAKVVLAADNWGDNTGYQFL-LDADHNTFGSVIP  1448

Qy  1460  PGGSSFPAGHNSAICVSSASHINFEG--PQNPD--------NYLVT--PELSLPGGGTLTFWV  1509
           .||:     ::|:    :||:: :    |:||         |: :       ||||:| | 
Db  1449  ATGPLFTGTASSNLYSA---NFEYLIPANADPVVTTQNIIVTGQGEVVIPGG---VYDY  1501

Qy  1510  CAQDANYASEHYAVVASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKT  1569
           | :: :  ::: :|: |:|| :  || ||
Db  1502  CITNPEPASGKMWI-AGDGGNQPARYDDFTFB------------------  1532

Qy  1570  VQLPAGTKYV-AFRHFGCTDFFWINLDDVVITSGNAPSYTTYIYRNNTQIASGVTETTYR  1628
                :|:|   ::: |::  |||||:|||  | ||||| |||:||: : :|:||||||||
Db  1533  ----AGKKYFTMRRAGMGDG------TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYR  1583

Qy  1629  DPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLVVGKITVTCQGEA  1688
           |    :  :| ||:|  ::||  : ::|::    ||||||||||:||||||||||||
Db  1584  DAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLIVVGKITVTCQGEA  1643

Qy  1689  MIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVDGKSYVEKLAVK  1732
           |||||||||||||||||||||||| |||||||||||||||||
Db  1644  MIYDMNGRRLAAGRNTVVYTAQGGYYAVMVVDGKSYVEKLAVK  1687


RESULT 15
AAR96033
ID   AAR96033 standard; protein; 1687 AA.
XX
AC   AAR96033;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagE.
XX
KW   Haemagglutinin; hagE; periodontal disease; vaccine; antibody.
```

```
XX
OS   Porphyromonas gingivalis; strain FDC381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   95WO-US016108.
PF   11-DEC-1995;
XX
PR   94US-0035485.
PR   09-DEC-1994;
XX
PA   (UFFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30656.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 138-143; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagE (AAR96033) was identified as the
CC   product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
CC   The haemagglutinin can be obtd. from transformed host cells and used as a
CC   vaccine to protect humans or animals against periodontal disease.
CC   Expression in Salmonella cells allows prodn. of a live vaccine. The
CC   haemagglutinin can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1687 AA;

Query Match       46.1%;   Score 4233;   DB 2;   Length 1687;
Best Local Similarity  51.5%;   Pred. No. 1.4e-271;
Matches  918;  Conservative 197;  Mismatches 445;  Indels 224;  Gaps 44;
```

```
Qy    47 FSFNEVELTKVETKGGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKS-- 104
Db    30 FRMDNLKFTEVQTPKGMAQVFTYTEGVNLSEKGMPTLPILSRSLAVSDTRMKVEVVSSK   89

Qy   105 FTEQVYSLNQYGSEKLMPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTM 164
Db    90 FIEKKNVL------IAPSKGMIMRNEDPKKIPYVY-GKSYSQNKFFPGEIATLDDPFIL  141

Qy   165 RGVRIAALTINPVQYDVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETA 221
Db   142 RDVRGQVVNFAPLQYNPVTKLRIYIEITVAVSETSEQGKNILNKKGTFAG-----FEDT  196

Qy   222 YKQLFNRDVVTDHGDLYNTPV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYT 275
Db   197 YKRMF-----MNYEPGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVVDWKNQRGLRTEVKVA 251

Qy   276 DEAEVGTTNASIKAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVT 329
Db   252 EDIASPVTANAIQQFVKQEYBKEGND------LTYVLLVGDHKDIPAKITPGIKSDQV-  303

Qy   330 DLYYSAVDGDYFPEMYTFRMSASSPEELTNIIDKVLMYBKATMPDKSYLEKVLLIAGADY 389
Db   304 --YGQIVGNDHYNEVFIGRFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASABG 361

Qy   390 SWNSQVGQPTIKY-GMQYYNQEHGYTDVNYLKAPYTGCY---------SHLNTGVSF    438
Db   362 GPSADNGEBSDICHENVIANLLTQYGYTKIIK------CYDPGVTPKNIIDAFNGGISL  413

Qy   439 ANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV-- 496
Db   414 VNYTGHGSETAWGTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQK 473

Qy   497 --KEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVN 554
Db   474 DGKPTGTVAIIASTINQSW-----------QDEMNEI-                       505

Qy   555 SIMWAGNLAATHAGNI---GNITHIGAHYYWEAXH--VLGDGSVMPYRAM           599
Db   506 ------LCEKHPNNIKRIFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLV  558

Qy   600 PKTNTYTLPASLPQNQASYSIQAS-AGSYYAISKDGVLYGTGVANASGVATVSMTKQITE 658
Db   559 PTKMQVTAPAQINLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTN 616

Qy   659 NGNYDVVITRSNYLPVIKQIQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK---KAE 714
Db   617 ESTLILTVVGYNKETVIKTINTNGEBPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTN 676

Qy   715 GSREVKRIGDGLFVTIEPAND-VRANEAKVVLAADNVWGDNTGYQPLLDADHNTFGSVIP 773
Db   677 TARSVDGIREIVLLSVSDAPELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQYGQVIP  736

Qy   774 A-TGPLFTG-TASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNP 831
Db   737 SDTHTLWPNCSVPANLF-APFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAP 795

Qy   832 EPASGKMWIAGDGGNQOPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTV 891
Db   796 Q-ANAKIWIAGQG---PTKEDDYVFEAGKKYHFLMKKMGSGDGTELTISBGGGSDYTYTV  851

Qy   892 YRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNL 951
Db   852 YRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNL 911

Qy   952 TGSSVGQKVTLKWDAPNGTPNPNPNPNPNPG---TTLSESFENGIPASWKTIDADGDGHG 1008
Db   912 TGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHG    971

Qy  1009 WKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDAN 1068
Db   972 WKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDAN 1031

Qy  1069 YASEHYAYASSTGNDASNFTNALLEETITAKGVRSPKAIRGRIOGTWRQKTVDLPAGTK  1128
Db  1032 YASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPEAIRGRIOGTWRQKTVDLPAGTK 1091

Qy  1129 YVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL 1188
Db  1092 YVAFRHFQSTDMFYIDLLDEVEIKANGKRADFTETFESSTHGAPAEWTTIDADGDGQGWL 1151

Qy  1189 CLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDH 1248
Db  1152 CLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDH 1211

Qy  1249 YAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTK  1308
Db  1212 YAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTK  1271

Qy  1309 YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYYVYRDGTKIKEGLTETTFEEDGVAT 1368
Db  1272 YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYYVYRDGTKIKEGLTETTFEEDGVAT 1331

Qy  1369 GNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE 1428
Db  1332 GNHEYCVEVKYTAGVSPKECVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTE 1389

Qy  1429 -------------------VLNEDFENGIPASWKTIDADGDGNNWTTTPP           1459
Db  1390 GSREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIP 1448

Qy  1460 PGGSSPAGHNSAICVSSASHINFEG--PQNPD-------NYLVT--PELSLPGGGTLTFWV 1509
Db  1449 ATGPLFTGTASSNLYSA---NFEYLIPANADPVTTQNIIVTGQGEVIPGG---VYDY    1501

Qy  1510 CAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKT 1569
Db  1502 CITNPEPASGKWMI-AGDGGNQPARYDDFTFE-----                        1532
```

```
Qy  1570 VQLPAGTKYV-AFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYR 1628
              |:||||||     |---|   :  :    ||||:|:::  |:|:|||||
Db  1533 ----AGKKYTFTMRRAGMGDG-----TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYR 1583

Qy  1629 DPDLATGFYTYGVKVVVPNGESAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEA 1688
              |  : .||::|:|  |     :||||||||||||||||||||||||
Db  1584 DAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQNPYTLTVVGKTITVTCQGEA 1643

Qy  1689 MIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVDGKSYVEKLAVK 1732
              ||||||| |||:||||||||||:|||||||||||||||||||
Db  1644 MIYDMNGGRLAAGRNTVVYTAQGGYYAVMVVDGKSYVEKLAVK 1687
```

Search completed: May 18, 2004, 11:32:46
Job time : 93.0274 secs

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 18, 2004, 11:29:09 ; Search time 31.8582 Seconds
                    (without alignments)
                    5229.534 Million cell updates/sec

Title:          US-08-353-485-10
Perfect score:  9179
Sequence:       1 MRKLLLLIAASLLGVGLYAQ.........HYAVMVVVDGKSYVEKLAVK 1732

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283366 seqs, 9191526 residues

Total number of hits satisfying chosen parameters:   283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      PIR_78:*
                1: pirl:*
                2: pir2:*
                3: pir3:*
                4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	9179	100.0	1732	2	T30836	lysine-specific cy
2	4270	46.5	1704	2	A55426	gingipain R (EC 3.
3	4217	45.9	2628	2	T28651	hemagglutinin A -
4	3808	41.5	1526	2	S49763	gingipain R (EC 3.
5	1222.5	13.3	991	2	I40229	arginyl endopeptid
6	322	3.5	5291	2	F90696	hypothetical prote
7	315	3.4	5188	2	B85547	probable RTX famil
8	313	3.4	1993	2	AF1450	probable peptidogl
9	300.5	3.3	2468	2	A83412	hypothetical prote
10	295.5	3.3	4199	2	S76412	hypothetical prote
11	280.5	3.1	3029	2	S76109	hypothetical prote
12	276.5	3.0	2044	2	AB1180	probable peptidogl
13	269	2.9	2167	2	AF1489	cell wall-associat
14	265	2.9	1283	2	T39174	hypothetical Serin
15	262	2.9	2059	2	D82671	surface protein XF
16	252	2.7	3083	2	AH2493	hypothetical prote
17	250.5	2.7	1200	2	T17404	hyalin - sea urchi
18	248	2.7	2554	2	AB3528	extracellular seri
19	247.5	2.7	1939	2	D97316	probable S-layer p
20	244	2.7	4936	2	AH2515	hypothetical prote
21	243	2.6	3972	2	S75251	hypothetical prote
22	242.5	2.6	2334	2	S32920	cell wall-associat
23	241	2.6	1904	2	T13256	tail-host specific
24	239	2.6	3624	2	AD0835	large repetitive p
25	238	2.6	1649	2	C86822	hypothetical prote
26	231.5	2.5	1461	2	B90696	hypothetical prote
27	230.5	2.5	1461	2	A85547	hypothetical prote
28	229.5	2.5	1483	2	C97012	probably celluloso
29	229	2.5	1873	2	T30944	surface protein pr
30	228.5	2.5	2130	2	AB0821	probable exported
31	227	2.5	3705	2	AD0123	probable autotrans
32	225	2.5	26926	1	I38344	titin, cardiac mus
33	224.5	2.4	1999	2	AB2018	hypothetical prote
34	224	2.4	2894	2	C64474	hypothetical prote
35	223	2.4	1541	2	A37023	IgA-specific metal
36	223	2.4	3262	2	AH2137	hypothetical prote
37	221.5	2.4	2508	2	S61441	surface-associated
38	221	2.4	2232	2	T34434	hypothetical prote
39	220.5	2.4	1902	2	S06997	lactocepin (EC 3.4
40	219.5	2.4	1441	2	A86685	prophage pil prote
41	219	2.4	1797	2	F69195	cell surface glyco
42	219	2.4	2902	2	C71953	toxin-like outer m
43	218	2.4	3283	2	AC1018	large repetitive p
44	216	2.4	1879	2	S74915	extracellular nucl
45	216	2.4	2893	2	A64556	toxin-like outer m

ALIGNMENTS

```
RESULT 1
T30836
lysine-specific cysteine proteinase porphypain (EC 3.4.22.-) - Porphyromonas gingivalis
N;Alternate names: lysine-specific cysteine proteinase 1, 60K
C;Species: Porphyromonas gingivalis
C;Date: 22-Oct-1999 #sequence revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30836; T30837; T30526; A53113
R;Barkocy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulske-Fox, A.; Lantz,
J. Bacteriol. 178, 2734-2741, 1996
A;Title: Analysis of the prtP gene encoding porphypain, a cysteine proteinase of Porphy
A;Reference number: Z20895; MUID:96213011; PMID:8631659
A;Accession: T30836
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1732 <BAR>
A;Cross-references: EMBL:U42210; NID:g1314325; PID:g1314326; PIDN:AAB06565.1
R;Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z20896
A;Accession: T30837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-795, 'I', 797-1389, 'N', 1391-1478, 'Y', 1480-1732 <SLA>
A;Cross-references: EMBL:U75366; NID:g182811; PID:g2182812; PIDN:AAB60809.1
R;Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A;Title: IS195, an insertion sequence-like element associated with protease genes in Por
A;Reference number: Z20844; MUID:98298016; PMID:9632563
A;Accession: T30526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1350, 'N', 1352-1363, 'Y', 1365-1447, 'H', 1449-1732 <LEW>
A;Cross-references: EMBL:AF017059; NID:g2738802; PID:g2738803; PIDN:AAC26523.1
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:9103245; PMID:8276827
A;Accession: A53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 229-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141690)
C;Genetics:
A;Gene: prtP; prtK
C;Keywords: cysteine proteinase; hydrolase

Query Match         100.0%;       Score 9179;   DB 2;   Length 1732;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 1732;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy        1 MRKLLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVBTK 60
```

```
Db
Qy          1  MRKLLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSRFNEVELTKVETK      60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db
Qy         61  GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL     120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61  GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL     120

Db
Qy        121  MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD     180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121  MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD     180

Db
Qy        181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT     240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT     240

Db
Qy        241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA     300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA     300

Db
Qy        301  ASAAPVFIALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPELTNI     360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301  ASAAPVFIALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPELTNI     360

Db
Qy        361  IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNY     420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361  IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNY     420

Db
Qy        421  LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA     480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421  LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA     480

Db
Qy        481  QFDYVYQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS     540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481  QFDYVYQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS     540

Db
Qy        541  YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP     600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541  YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP     600

Db
Qy        601  KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG     660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601  KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG     660

Db
Qy        661  NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK     720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661  NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK     720

Db
Qy        721  RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT     780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721  RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT     780

Db
Qy        781  GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI     840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781  GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI     840

Db
Qy        841  AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVVRDGTKIKE     900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841  AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVVRDGTKIKE     900

Db
Qy        901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV     960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV     960

Db
Qy        961  TLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY    1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961  TLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY    1020

Db
Qy       1021  NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS    1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021  NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS    1080

Db
Qy       1081  TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM    1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db
Qy       1081  TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM   1140

Db
Qy       1141  FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA   1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141  FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA   1200

Db
Qy       1201  HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA   1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201  HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA   1260

Db
Qy       1261  GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL   1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261  GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL   1320

Db
Qy       1321  NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT   1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321  NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT   1380

Db
Qy       1381  AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA   1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381  AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA   1440

Db
Qy       1441  SWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP   1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441  SWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP   1500

Db
Qy       1501  GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR   1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501  GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR   1560

Db
Qy       1561  AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIAS   1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561  AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIAS   1620

Db
Qy       1621  GVTETTYRDPDLIATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTTVGKTI   1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621  GVTETTYRDPDLIATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTTVGKTI   1680

Db
Qy       1681  TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGHYAVMVVVDGKSYVEKLAVK          1732
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681  TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGHYAVMVVVDGKSYVEKLAVK          1732
```

RESULT 2
A55426
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
N;Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP; F
C;Species: Porphyromonas gingivalis
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
C;Accession: A55426; D53113
R;Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr, I
J. Biol. Chem. 270, 1007-1010, 1995
A;Title: Molecular cloning and structural characterization of the Arg-gingipain protein
A;Reference number: A55426; MUID:95138080; PMID:7836351
A;Accession: A55426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1704 <PAV>
A;Cross-references: GB:U15282; NID:g557067; PIDN:AAA69539.1; PID:g557068
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: D53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 228-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141694)
C;Keywords: cysteine proteinase; hydrolase

Query Match   46.5%; Score 4270; DB 2; Length 1704;
Best Local Similarity   50.9%; Pred. No. 1.2e-228;

```
Qy  1023 NGCVYSBSFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQODANYASEHYAVYASSTG 1082
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1003 NGCVYSBSFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQODANYASEHYAVYASSTG 1062

Qy  1083 NDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKVVAFRHFQSTDMFY 1142
         |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  1063 NDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKVVAFRHFQSTDMFY 1122

Qy  1143 IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHG 1202
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1123 IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHG 1182

Qy  1203 GSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGD 1262
         :||| :||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1183 GTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGD 1242

Qy  1263 FTVVFEETPNGINKGGARFGLSTEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNY 1322
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1243 FTVVFEETPNGINKGGARFGLSTEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNY 1302

Qy  1323 ILLDDIQFTMGGSPTPTDYTYTVYYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG 1382
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1303 ILLDDIQFTMGGSPTPTDYTYTVYYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG 1362

Qy  1383 VSPKKCVDVTVNSTQFNFVQNLTAEQAPNSMDAILKWNAPASKRAE------------ 1428
         ||||:||:||:|:|||:|| ||: :                                :
Db  1363 VSPKECVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTEGSREVKRIGDGLFV 1420

Qy  1429 ------------VLNEDFENGIPASWKTIDADGDGNNWTTPPPGGSSFAGHNSAIC 1473
                     |::
Db  1421 TIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSNL 1479

Qy  1474 VSSASHINFEG--PQNPD------NYLVT--PELSLPGGGTLTFWVCAQDANYASEHYAV 1523
         :  |||      |          |||    :||:|||| ||||||
Db  1480 YSA-----NFEYLIPANADPVVTTQNIIVTGQGEVVIPGG---VDYCITNPEPASGKMWI 1532

Qy  1524 YASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-AFR 1582
         |::                                              ||||||||
Db  1533 -AGDGGNQPARYDDFTFE--------------------------------AGKKYTFTMR 1559

Qy  1583 HFGCTDFFWINLDDVVITSGNAPSYTYTYRNNTQIASGVTETTYRDPDLIATGFYTGYVK 1642
         |  :                 ||SYTYTY|RDGTKIKEGLTETTYRDAGMSAQSHEYCVE
Db  1560 RAGMGDG---------TDMEVEDDSPASYTYTYYRDGTKIKEGLTETTYRDAGMSAQSHEYCVE 1614

Qy  1643 VVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTIIVTCQGEAMIYDMNGRLAAGR 1702
         | | .   :    :|| |||||||||||||||||||||||||||||||||||||||||
Db  1615 VKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTIIVTCQGEAMIYDMNGRRLAAGR 1674

Qy  1703 NTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
         ||||||||||.||||||||||||||||||
Db  1675 NTVVYTAQGGYYAVMVVVDGKSYVEKLAVK 1704
```

```
RESULT 3
T28651
hemagglutinin A - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C;Accession: T28651
R;Han, N.; Whitlock, J.; Progulske-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A;Title: The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381 contains four 1
A;Reference number: Z20494; MUID:97047672; PMID:8926061
A;Accession: T28651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2628 <HAN>
A;Cross-references: EMBL:U41807; NID:g1552410; PID:g1469916; PIDN:AAB17128.1
C;Genetics:
A;Gene: hagA

Query Match    45.9%;    Score 4217;    DB 2;    Length 2628;
```

```
Matches 932;  Conservative 204;  Mismatches 456;  Indels 238;  Gaps 47;

Qy    6 LLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELIKVETKGGTFA 65
        : ::|| ||| ||||:::| |                  | :|||||||: ||
Db   10 IALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGI-- 65

Qy   66 SVSIPGAFPT------GEVGSPEVPAVRKLIAVPVGATPVVRVKS--FTEQVYSLNQYG 116
        |: |                   : :  |  :    :                 |
Db   66 --GQVPYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL----- 116

Qy  117 SEKLMPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINP 176
        : ::         |||: :::| |||  :::                     :
Db  117 ---IAPSKGMIMRNEDPKKIPYVY-GKSYSQNKFFPGEIATLDDPFILRDVRGQVNFAP 172

Qy  177 VQYDVVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTD 233
        :|| ||:| :|  : |||:   || :|  |  :  |  :  |||:|| ||::
Db  173 LQYNPVTKTLRIYTEITVAVSETSEQGKNILNKKGTFAG------FEDTYKRMF------MN 222

Qy  234 HGDLYNTPV------RMLVVAGAKFKBALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASI 287
        : | |||       |::|:| || |    |  :  :: | ::        :|  : |||
Db  223 YEPGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI 282

Qy  288 KAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKVTDLYYSAVDGDYF 341
        :  : ::|    | ||  |:| ||  ||| :| |||    | :  |:       ::| |
Db  283 QQFVKQEYEKEGND------LTYVLLVGDHKDIPAKITPGIKSDQV--YCQIVGNDHY 332

Qy  342 PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK 401
        :|:  |   |: |::|  :| :|: |:|  : ||||||:||::|| | | ||  | :|:|
Db  333 NEVFIGRFSCESKEDLKTQIDRTIHYERNITEDKWLGQALCIASAEGGPSADNGESDIQ 392

Qy  402 Y-GMQYYYNQEHGYTDVVNYLKAPYTGCY------SHLNTGVSFANYTAHGSETAW 450
        : |: | :||:|||  :|||| |       |  |||:| :|::|||||||||||||
Db  393 HENVIANLLFQVGYTKLIK-------CYDPGVTPKNIIDAFNGGISLVNYTGHGSETAW 444

Qy  451 ADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV---KEKGAYAYIG 506
        |     ||:|  |||:::: : | ::    ||| :|| |||:: |    :| |   :|
Db  445 GTSHFGTTHVKQLTNSNQLPFIPDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIA 504

Qy  507 SSPNSYWGEDYYWSVGANAVFGVQPTFRGTSMGSYDATFPLEDSYNTVNSIMWAGNLAATH 566
        |: |||    ||                        |::  :||:  :
Db  505 STINQSW----ASPMRG------------------QDEMNEI------LCEKH 529

Qy  567 AGNI----GNITHIGAHYYWEAYH--------VLGDGSVMPYRAMPKTNTYTLPASL 611
        | |:    |::  :|  :  :          ::|||  :  |  :| :::| :|||
Db  530 PNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQI 589

Qy  612 PQNQASYSIQAS-AGSVYAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN 670
        . |:||:|| |  ||:  |||:|::   |  ||  | |: ::  :::||| |::||: ||
Db  590 NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTILILTVVGYN 647

Qy  671 YLPVIKQIQV-GEPSPYQPVSNLTATTGQKVTLKWEAPSAK--KAEGSREVKRIGDGL 726
        : |:||   : ||| |||||||||||||||||||||:||| |  :|  | ::   ||:
Db  648 KETVIKTINTNGEPNPYQPVSNLTATTGQKVTLKWDAPSTKTNATTNARSVDGIRELV 707

Qy  727 FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA 783
        : :: |:   : |  ||:||||| | | :|||:|||||||:  | ||  ||| ||| ||
Db  708 LLSVSDAPELLRSGQAEIVLEAHDVWNDSGYQILLDADHDQYGQVIPSDTHTLWPNCSV 767

Qy  784 SSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGD 843
        .:||:  ||| |||:||||  | :|:|  |  :  |||||||:||||||||||||||||
Db  768 PANLF-APFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQ 825

Qy  844 GGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLT 903
        |   |::|||| ||||||||| ::  ||||||::  :| | ||||||||||||||||||:
Db  826 G---PTKEDDYVFEAGKKYTFLMKKMGSGDGTELTISEGGSDYTYTVYRDGTKIKEGLT 882

Qy  904 ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK 963
        |||||||||:|||||||||||||||||||||||||||||||||||||||||:|||||||
Db  883 ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK 942

Qy  964 WDAPNGTPNPNPNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNS 1022
        |||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db  943 WDAPNGTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNS 1002
```

Best Local Similarity   51.0%;   Pred. No. 2.2e-225;
Matches 958;   Conservative 114;   Mismatches 363;   Indels 442;   Gaps 53;

```
Qy    179 YDVVANQLKVRNNIEIEVSFQ---------GADEVATQRLYDASFSPYFETAYKQLFNRDVVT 232
                :|:: :|::||                |:  |: :: ||
Db    871 YTVVRDGTKIKEGL-TETTFEEDGVATGNHEYCVEVKYTAGVSP-----KECVNVTV-     921

Qy    233 DHGDLYNTPVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIH  292
          |:|:              ||:  |                :|| |:::: |
Db    922 --DPVQFNPVQNLTGSAVGQKVTLK-----WDAPNG---TPNPNPNPGTTTLS          966

Qy    293 KKYNDGLAAS---------------AAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAV   336
          :|:: :|:                  ||        ||:|
Db    967 ESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSAS-------YINF 1017

Qy    337 DGDYPPEMYTFRMSASSP--------------BELTNIIDKVLMYEKATMPDKS------YLEK  380
           |: :         |:                :||: |||: ::|     :||
Db   1018 EGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEE 1077

Qy    381 VL---------LIAG--ADYSWNSQVGQPTIKYGMQYYNQEHGYTDVY------------- 418
          ||         |:|    |:              |: :::|| |:
Db   1078 VLTAKTVVTAPEAIRGTRVQGTWYQKTVQ--LPAGTKYVAFRHFGCTDFFWINLDDVEIK 1135

Qy    419 -NYILKAPYTGCYSHLNTGVSPANYT---AHGSETAWADPLLTTSQLKALINKDKYFLAIG  474
             ::|| |:  |    |:|||            ::  ||  :||  :|
Db   1136 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWL--CLSSGQLGWLT-----AHG    1186

Qy    475 NCCITAQFDY-----VQP----CFGEVITRVKEKGAYAYIGSSPNSYWGEDYWSVGANAV  526
          |:  ||          |        ||:|  ||:  : ||:  |::  |
Db   1187 GTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDH--------------- 1232

Qy    527 FGVQPTFEGTSMGSYDATFLEDSYNTVN--------------SIMWAGNLAATHAGN    569
          :|            |:   ||:                    |:  |:: |
Db   1233 YAVMISKTGTNAGDFTVVF-EETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGT 1291

Qy    570 IGNITHIGAHYY-----WEAYHVLGDGSVMPYRAMPKTNTYTLPASLPQNQASYSIQASAGSY  627
          :|||     ||:     :|      |:   ||  |     |:|
Db   1292 ---KYVAFRHYNCSDLNYILLDDIQFTMGGSPTFPTDYT------------------    1326

Qy    628 VAISKDQVLYGTGVANAS------GVATVSMTKQITENGNYDVVITRSNYLPVIKQIVG--  681
           |:|       |||        |:  |::   |  |:|  ||:  ::|| |:|
Db   1327 YTVVRDGTKIKEGLTETTFEEDGVAT----------GNHEYCVEVKYTAGVSPKECVNVT 1376

Qy    682 -EPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPANDVRANE   740
           |         ||||:|:||                  | |
Db   1377 VDPVQFNPVQNLTGSAVGQKVTLKWDAPNG------------------TPNPNPNPNP   1416

Qy    741 AKVVLAADNWGDNTGYQFL-LDADHNTFGSVIPATGPLFTG-------TASSNLYSANFE-  793
          |:             |:|  |||:     |::|   | ||        :||| | |:|
Db   1417 GTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASSASY-INFEG 1475

Qy    794 -----YLVP---------------------------ANA--DPVV                 804
               ||:|                            ||    |::
Db   1476 PQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVL 1535

Qy    805 TTQNII------------VTG---QGEVIPGG                              822
          ||:::              :|    ||:|:||
Db   1536 TAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGK 1595

Qy    823 --------------VYDYCITNPEPA-------------SGKM-WIAGDGGN           846
                        ::|                     |:|  ::|  |::
Db   1596 RADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTNVVASFSWNGM 1655

Qy    847 --QPARY------------------------------DDFT--FE--------        857
            |||||                              ||||  ||
Db   1656 ALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGIN 1715

Qy    858 -----------AGKKYTFTMRRAGMGDGTDMEVED---------                 881
                      ||||||  | :|  ||  |: |:|
Db   1716 KGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYV-AFRHYNCSDLNYILLDDIQFTMGG 1774

Qy    882 -DSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVE  940
            ||| ||||||||||||||| | ||||||||| |||||||||||||||||
Db   1775 SPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVD 1834
```

```
Db   1775 SPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVD 1834
               ||||||||||||||||| | ||||||||| ||||||||||||||||||
Qy    941 GSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTI 1000

Qy    941 GSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTI 1000
               | :||||||| | |||||||||||||||| |||   | | | |||||||||||||||
Db   1835 -PVQFNPVQNLTGAVGQKVTLKWDAPNGTPNPNPNPNPGT-TTLSESFENGIPASWKTI 1892

Qy   1001 DADGDGHGWKPGNAPG---IAGTNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGK 1057
          |||||||  | |||     |||||   |:  || :: ::::: :|||::|: |||||||
Db   1893 DADGDGNNWTTTPPPGGTSFAGHNSAICVSSASY-INFEGPQNPDNYLVTPELSLPGGGT 1951

Qy   1058 LTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKG-VRSPKAIRG-RIQGT 1115
          ||||||||||||||||||||||||||||||||||||| |:|:| |:|| ||||  ||:
Db   1952 LTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGI 2011

Qy   1116 WRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEW 1175
          | |||||||||||||||||| :||  |: ||||||||||||||||||||||||||||||
Db   2012 WYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIKANGKRADFTETFESSTHGEAPAEW 2071

Qy   1176 TTIDADGDGQWICLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVK  1235
          ||||||||||| ||||||| ||||||| |||: || ||||||||||||||||||||||
Db   2072 TTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVK  2131

Qy   1236 YYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSV 1295
          ||||||||||||||||||||||| ||||||||||||||||||||||||||| ||||||
Db   2132 YYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEADGAKPQSV 2191

Qy   1296 WIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEG 1355
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2192 WIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEG 2251

Qy   1356 LTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDTVNSTQFNPVQNLTAEQAPNSMDA 1415
          |||||||||||||||||||||||||||||: | ||:|: |||||||||||||||||||
Db   2252 LTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTINPFTQFNPVQNLTAEQAPNSMDA 2311

Qy   1416 ILKWNAPASKRAEVLNEDFENGIPASWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVS 1475
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   2312 ILKWNAPASKRAEVLNEDFENGIPSSWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVS 2371

Qy   1476 SASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 1535
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2372 SASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 2431

Qy   1536 ANALLEEVLTAKTVVTAPEAIRGTRAOGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 1595
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   2432 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 2491

Qy   1596 DVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLLATGFVTYGVKVVYPNGESAIETA 1655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2492 DVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLLATGFVTYGVKVVYPNGESAIETA 2551

Qy   1656 TLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYA 1715
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2552 TLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYA 2611

Qy   1716 VMVVVDGKSYVEKLAVK 1732
          |||||||||||||||||
Db   2612 VMVVVDGKSYVEKLAVK 2628
```

RESULT 4
S49763
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis (fragment)
C;Species: Porphyromonas gingivalis
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997
C;Accession: S49763
R;Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.
submitted to the EMBL Data Library, November 1994
A;Description: Cloning, sequence analysis and expression in Escherichia coli of prpR1 of
A;Reference number: S49763
A;Accession: S49763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <ADU>

```
A;Cross-references: EMBL;X82680
C;Genetics:
A;Gene: prpR1
C;Keywords: cysteine proteinase; hydrolase

Query Match       41.5%;   Score 3808;  DB 2;  Length 1526;
Best Local Similarity  51.7%;  Pred. No. 4.7e-203;
Matches 817;  Conservative 183;  Mismatches 403;  Indels 176;  Gaps 38;
```

```
Qy  911  GVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGT  970
         ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db  890  GVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGT  949

Qy  971  PNPNPNPNPNPG---TTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGCVY   1027
         ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db  950  PNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGCVY     1009

Qy 1028  SESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASN  1087
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1010  SESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASN  1069

Qy 1088  FTNALLEETTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDE   1147
         ||||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||
Db 1070  FTNALLEETTAKGVRSPEAIRGRIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDE   1129

Qy 1148  VEIKANGKRADFTETFBSSTHGEAPAEWTTIDADGDGQGWlCLSSGQLDWlTAHGGSNVV  1207
         |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db 1130  VEIKANGKRADFTETFBSSTHGEATAEWTTIDADGDGQGWlCLSSGQLDWlTAHGGTNVV  1189

Qy 1208  SSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVF   1267
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1190  SSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVF   1249

Qy 1268  EETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDD  1327
         ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db 1250  EETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDD  1309

Qy 1328  IQFTMGGSPTPTDYYTVYRDGTKIKEGLTETTFEBDGVATGNHEYCVEVKYTAGVSPKK   1387
         ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db 1310  IQFTMGGSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKK   1369

Qy 1388  CVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE-----               1428
         || |||||||||||| ||  :                                        
Db 1370  CVNVTVNSTQFNPVKNLKAQ--PDGDDVVLKWEAPSAKKTBGSREVKRIGDGLFVTIEPA  1427

Qy 1429  ------VLNEDFENGIPASWKTIDADGDGNWWTTTPPPGGSSFAGHNSAICVSSAS      1478
               ||   :        :   ::  :                                  
Db 1428  NDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSDLYSA--  1484

Qy 1479  HINFEG--PQNPDNYLVTP                                           1495
            ::    :                                                    
Db 1485  --NFEYLIPANADRNVRPP                                           1501
```

```
RESULT 5
I40229
arginyl endopeptidase - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40229
R;Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
A;Title: Structural characterization of argingipain, a novel arginine-specific cysteine
Arch. Biochem. Biophys. 316, 917-925, 1995
A;Reference number: I40229; MUID:95168884; PMID:7864651
A;Accession: I40229
A;Status: preliminary;  translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-991 <RES>
A;Cross-references: GB:D26470; NID:g927644; PIDN:BAA05484.1; PID:g927645

Query Match       13.3%;   Score 1222.5;  DB 2;  Length 991;
Best Local Similarity  31.8%;  Pred. No. 8.7e-60;
Matches 322;  Conservative 174;  Mismatches 381;  Indels 135;  Gaps 32;
```

```
Qy    6  LLIAASLLGVGLYAQSAKIKLDAPTTRTTCNNSFKQFDASFSFNEVELTKVETKGGTFA   65
         : :: ||||  : :|| :: ||||||  ::|:||                           
Db   10  IALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGMAQ   67

Qy   66  SVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKS--FTEQVYSLNQYGSEKLMPH  123
         |  ||||||||| ||||| ||  :|  :|  :   ::                        
Db   68  VPTYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVSSKFIEKKNVL-----IAPS    120

Qy  124  QPSMSKSDDPEKVFPVYNAAAYARKGFVGQBLTQVEMLGTMRGVRIAALTINPVQYDVVA  183
         :   :    ::||: :||  ::   ||:||:                              
Db  121  KGMIMRNEDPKKIPYVY-GKSYSQNKFFPGEIATLDDPFILRDVRGQVVNFAPLQYNPVT  179

Qy  184  NQLKVRNNIEI--EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT   240
         : || :      :: :                            ::              
Db  180  KTLRIYTEITTVAVSETSEQGKNILNKKGTFAG------FEDTYKRMF            229

Qy  241  PV------RMLVVAGAKFKEALKPWLTWKAQKGFVLDVHYTDEAEVGTTNASIKAFIHKK  294
         ||      ||:|: |||  ||  :                   : :              
Db  230  PVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIIASPVTANAIQQFVKQE 289

Qy  295  Y---NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYSAVDGDYFPEMYTFR   348
         |    :  :   | ||:|||||||::|:    :||  || :                   
Db  290  YEKEGND------LTYVLLIGDHKDIPAKITPGIKSDQV---YGQIVGNDHYNEVFIGR   339

Qy  349  MSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKY-GMQYY  407
         :  :    |  :|:  : :: :  :  :                                 
Db  340  FSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQHENVIAN  399

Qy  408  YNQEHGYTDVYNYLKAPYTGCY----------SHLNTGVSFANYTAHGSETAWADPLLLT  457
          :  :||| |:| || |                ::  |  ::||| ||||||| |      
Db  400  LLITQYGYTKIIK-------CYDPGVTPKNIIDAFNGGISLANYTGHGSETAWGTSHFGT  451

Qy  458  SQLKALTNKDKYFLAIGNCCITAQPDYVQPCFGEVITRV---KEKGAYAYIGSSPNSW    513
         : :| |||::: : | : || : :| | :| |||::         :    ||:: ::|: 
Db  452  THVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIASTINQSW  511

Qy  514  GEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNI--  570
           :       : :||                      :  ::                  
Db  512  --ASPMRG--------------------------QEMNEI------LCEKHPNNIKRT   536

Qy  571  -GNITHIGAHYYWEAYH------------VLGDGSVMPYRAMPKINTYTLPASLFQNQASY 618
          ||:| | |  :                 |:|| :: :    :                  
Db  537  FGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPFTKMQVTAPAQINLTDASV 596

Qy  619  SIQAS-AGSVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIKQ   677
         ::  : ||: |||  | :     ::  |:|:|::  |   : :                  
Db  597  NVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGYNKETVIKT  654

Qy  678  IQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK---KAEGSREVKRIGDGLFVTIEPA  733
         | : ||| |||||||||||||||||||||:||| |   :|   ::| : || ||| :   
Db  655  INTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDA  714

Qy  734  ND-VRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TASSNLYSA   790
          : ||::|| :: |||||:||:|||:|||||||: :| |||  || | | |||:| ||  
Db  715  PELLLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQYGQVIPSDTHTLWPNCSVPANLF-A 773

Qy  791  NFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVPDYCITNPEPASGKMWIAGDGNQPAR  850
         :||: |||:||   :  :|::: :|||:||:| :   ::  || |  | ||||:|      
Db  774  PFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQG---PTK  829

Qy  851  YDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEED  910
          ||:|||||||: | :::|| ||||::  :  |||:||||||||||||||||||||||||| 
Db  830  EDDYVFEAGKKYHFLMKKMGSGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEED  889
```

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5291 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33965.1; PID:g13360000; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0542

Query Match      3.5%;     Score 322;  DB 2;   Length 5291;
Best Local Similarity   19.6%;  Pred. No. 1.3e-08;
Matches  365;  Conservative 228;  Mismatches 697;  Indels 572;  Gaps  86;

Qy   163  TMRGVRIAAL----------TINPVQYDVVANQLKVRNNIEIEVSFQGA-DEVATQRLYDASFS  215
          : :  :::                 ::.::  :::  :.:::::.  :.:
Db  1696  TVPASRVSALGEATYTVTAAATDADGNSGSASHNVQVNTALPGVTINVVATDDIINAA--    1753

Qy   216  PYFETAYKQLFNRDVY-TDHGDLYNTPVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHY    274
Db  1754  ---EAGVEQTISGQVTGAAAGDTVT-----VTLGGATYTATVQANLSWS----VDVPA    1799

Qy   275  TDEAEVGTTNASIKAFIHKKYNDGLAASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYS    334
Db  1800  SALQELGNGELTISASVTNS----------------VGNT-----GNGTREIT-----    1831

Qy   335  AVDGDYFPEMYTFRMSASSPEELTNIID-KVLMYEKA----------TMPDKSYLEK     380
Db  1832  -IDAN----LPGLRVDTVAGDDVVNIIEHGQALVITGSSSGLAAGSNVTLTINGQTYVAA  1886

Qy   381  VLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNYLKAPYTGCYSHLNTGVSFAN  440
Db  1887  VL----ADGTWS--VGVPAV--------------------DVSAWPAGSVT------    1916
                                                                IAASG

Qy   441  YTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRVKEKG  500
Db  1917  STSAGNPVSVTHPV--TVDLSA------------------VAVSINATTA------     1956
                                                                DDVINAAEKG

Qy   501  AYAYIGSSPN--------------SYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLED--  548
Db  1957  AALTLSGSTSGVEAGQTVTVTFGGKTYSATVAANGSW------STSVPAADMAALRDGD   2009

Qy   549  -----SYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMPKTN   603
Db  2010  ASAQASVSNVN-----GNSATT-------------------TH   2028

Qy   604  TYTLPASLPQNQASYIQASAGSYVAISKDG---VLYGTGVANASGVATVSM-----TK   654
Db  2029  AYSVDASAP------TVTINTIAGDDILNAAEAGAALTITGSSTAEAGQTVTVTLNGTNYTG  2084

Qy   655  QITENGNYDV----------VITRSNY-LPVIKQIQVGEPSPYQPVSNLTATTQGGQKVTLKW  705
Db  2085  TVQTDGSWSVSVPSADLSTLIASNYTVNAAVSDKAGNPASVN--HNLTVDTSVPVVTINT   2142

Qy   706  EA----------PSAKKAEGSREVKRIGDGLFVTIEPA   733
Db  2143  VAGDDVINATEHAQAQIISGSATGAATGSTVTVTIGINTFTTVLDASGNWSVGVPASVVS   2202

Qy   734  ---------NDVRANEAKVVLAADNVWGDNTGYQFLLDADHN-----   766
Db  2203  ALANGTVTINASVTDAGGNSGSGATHQVTVNTGLPTTIFNAISGDN-----ILNADEKGQP  2257

Qy   767  ---TFGSVIPATGPLFTGTASSNLYSA-----------NFEYLVPANADPVTTQNIIV-----  811
Db  2258  LTISGGGTGLATGAQVTVLNGHNYSATTDASGNWTLTVPVSDLAALGQANYTVSASATS    2317

Qy   812  ----TGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTEAGKKYTFT   865
Db  2318  AAGNTASSQANLLVDSGLPDVTINT----------VAGDDIINAA--------EAGADQTIS  2361

Qy   866  --MRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEV  923
Db  2362  GVVTRAAAGDTVTVTLGGN---TYTATVQ---SNLSWSVSVPTADLQALGNGDLTITASV   2415

Qy   924  KYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSV------GQKVTLKWDAPNGTPNPNPNP   977
```

```
Qy   124  QPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYDVVA  183
          : :::  :  :::         :::::   :.::::::.         :::
Db   121  KGMIMRNEDPKKIPYVY-GKSYSQSNKFPPGBIATLDDPFIIRDVRGQVVNFAPLQYNPVT  179

Qy   184  NQLKVRNNIEI---EVSFQGAEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT   240
Db   180  KTLRIYTETTVAVSETSEQGKNILNKKGTFAG------FEDTYKRMF------MNYEPGRYT  229

Qy   241  PV------RMLVVAGAKFKEALKPWLTWKAQKGFVLDVHYTDEAEVGTTNASIKAFIHKK  294
Db   230  PVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAIQQFVKQE   289

Qy   295  Y-----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYYSAVDGDYFFPEMYTFR  348
Db   290  YEKEGND--------LTYVLLVGDHKDIPAKITPGIKSDQV---YQOIVGNDHYNEVFIGR  339

Qy   349  MSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKY-GMQYY  407
Db   340  FSCESKEDLKTQIDRTIHYERNITTEDKNILGQALCIASAEGGPSADNGESDIQHENVIAN  399

Qy   408  YNQEHGYTDVYNYLKAPYTGCY----------SHLNTGVSFANYTAHGSETAWADPLLT   457
Db   400  LLTQYGYTKIIK--------CYDPGVTPKNIIDAFNGGISLVNYTGHGSETAWGTSHFGT   451

Qy   458  SQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV----KEKGAYAYIGSSPNSYW  513
Db   452  THVKQLTNSNQLFFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIASTINQSW  511

Qy   514  GEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNI--   570
Db   512  ---------ASPMRG---------QDEMNEI-------LCEKHPNNIKRT   536

Qy   571  -GNITHIGAHYWEAYH--------VLGDGSVMPYRAMPKTNTYTLPASLPQNQASY    618
Db   537  FGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQINLTDASV  596

Qy   619  SIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIKQ  677
Db   597  NVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLITVVGNKETVIKT  654

Qy   678  IQV-GEPSPYQPVSNLTATTGQGKVTLKWEAPSAK---KAEGSREVKRIGDGLFVTIEPA  733
Db   655  INTNGEPNPYQPVSNLTATTGQGKVTLKWEAPSAK---KAEGSREVKRIGDGLFVTIEPA  714

Qy   734  ND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TASSNLYSA  790
Db   715  PELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQYGQVIPSDTHTLWPNCSVPANLF-A  773

Qy   791  NFEYLVPANADPVTTQNIIVTGQGEVIPGGVYDYCITNPEPASGKMWIAGDGGNQPAR   850
Db   774  PFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGGG--PTK  829

Qy   851  YDDFTEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEED  910
Db   830  EDDYVFEAGKKYHFLMKKMGSGDGTELTISEGGSGDYTYTVYRDGTKIKEGLTETTYRDA  889

Qy   911  GVAAGNHEYCVEVKYTAGVSPKVCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTL  962
Db   890  GMSAQSHEYCVEVKYAAGVSPKVCVCVDYIPDGVADVTAQKPYTLTVVGKTITV   941
```

RESULT 6

```
F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: F90696
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90696
```

C;Accession: B85547
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dinalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85547
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5188 <STO>
A;Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:Z06
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0615

Query Match   3.4%;   Score 315; DB 2; Length 5188;
Best Local Similarity 19.5%; Pred. No. 3e-08;
Matches 363; Conservative 228; Mismatches 700; Indels 570; Gaps 85;

```
Qy   163 TMRGVRIAAL------TINPVQYDVVANQLKVRNNIELEVSFQGA--DEVATQRLYDASFS 215
             :  ::|::|             :::::  :: ::    :::
Db  1696 TVPASRVSALGEATYTVTAAATDADGNSGSASHNVQVNTALPGVTINVVATDDIINAA-- 1753

Qy   216 PYFETAYKQLFNRDV--TDHGDLYNTPVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHY 274
                                :    ||   :  :|
Db  1754 ---EAGVEQTISGQVTGAAAGDTVT------VTLGGATYTATVQANLSWS------VDVPA 1799

Qy   275 TDEAEVGTTNASIKAFIHKKYNDGLAASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYS 334
             :  |  :                    ||:
Db  1800 SALQELGNGELTISAGVTNS------------------VGNT----GNGTREIT 1831

Qy   335 AVDGDYFPEMYTFRMSASSPEELTNIID--KVLMYEKA----TMPDKSYLEK 380
                           :  :  :::: :             :  ||
Db  1832 -IDAN---LPGLRVDTVAGDDVVNIIEHGQALVITGSSSGLAAGSNVTLIINGQTVVAA 1886

Qy   381 VLLIAGADYSWNSQVGQOPTIKYGMQYYYNQEHGYTDVYNYLKAPYTGCYSHLNTGVSFAN 440
                                                         ||
Db  1887 VL----ADGTWS--VGVPAV------DVSAWPAGSVT------IAASG 1916

Qy   441 YTAHGSETAWADPLLITSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRVKEKG 500
              :   :        |:                              :  |
Db  1917 STSAGNPVSVTHPV--TVDLSA------VAVSINAITA----DDVINAAEKG 1956

Qy   501 AYAYIGSSPN------------------SYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLED- 548
             :  ::                                  :  |:
Db  1957 AALTLSGSTGVEAGQTVTVTFGGKTYSATVAANGSW------STSVPAADMAALRDGD 2009

Qy   549 -------SYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTN 603
                 :  :::                                            --TH
Db  2010 ASAQASVSNVN------GNSATT 2028

Qy   604 TYTLPASLPQNQASYSIQASAGSYVAISKDG---VLYGTGVANASGVATVSM----TK 654
             :  :                      :::|                 :::|:
Db  2029 AYSVDASAP---TVTINTIAGDDILNAAEAGAALTITGSSTAEAGTVTVTLNGTNYTG 2084

Qy   655 QITENGNYDV------VITRSNY-LPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKW 705
                    :  :::  :                                  :   |||
Db  2085 TVQTDGSWSVSVPSADLSTLTASXYTVNAAVSDKAGNPASVN--HNLTVDTSVPVVTINT 2142

Qy   706 EA------PSAKKAEGSREVKRIGDGLFVTIEPA--------- 733
             ||
Db  2143 VAGDDVINATEHAQAQIISGSATGAATGSTVTVTIGTNTFTTVLDASGNWSVGVPASVVS 2202

Qy   734 ------NDVRANEAKVVLAADNVWGDNTGYQFLLDADHN--- 766
                   :   :::|                       ||::
Db  2203 ALANGTVTINASVTDAGGNSGSATHQVTVNTGLPTIIFNAISGDN---- 2257

Qy   767 -----TFGSVIPATGPLFTGTASSNLYSA-------NFEYLVPANADPVVTQNIIV----- 811
             :   ::                                   |||
Db  2258 LRTISGGSTGLATGAQVTVTLNGHNYSATTDASGNWTLTVPVSDLAALGQANYVSASATS 2317

Qy   812 -----TGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGNGNQPARYDFTPFAGKKYTFT 865
```

```
Db  2416 TNANGNTGSGGTRDITIDANLPGLRVDTVAGDDIVNSIEHGQALVI------TGGSS 2465
              :  :||:::: ::                                       ||
Qy   978 NFNPGTTLSESFENGIPASWKTIDADGDGHGWKPG-NAPGIAGYNSNGC--VYSESFGL 1033
                     :|:                                :  |:  |:
Db  2466 GLNAGAVLTVTI-NSVAYS-ATVQADG---SWSVGIPAANVSAWPAGPLTVEVDGQSSAN 2520
Qy  1034 GGIGVLTPDNYLITPALDLPNGGKLITFWVCAQDANYASE---HYAVYASSTGNDASNFTN 1090
             ||                              :::|:::
Db  2521 NPVSVSHPFTVDLTAVAISIN-----TVASDDVINAAEKGTNLTLSGSTSGIESGQTV- 2573
Qy  1091 ALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1150
               :|::                    :|::|||                :
Db  2574 -----TVTFGGKTYTASVAA--NGSW---SVNVPA--------ADLATLPEGAANV 2611
Qy  1151 KANGKRADFTETFESSTHG---EAPAEWTTIDADGDG------QGWLCLSSG 1193
              : :|                  :::                 |:
Db  2612 QASVSSASGNSA--SATHAYSVDASAPTLTINTIASDDILNAAEAGSPLTISGTSTAETG 2669
Qy  1194 QLDWLTAHGGS-----NVVSSFSWNGM-------ALNPDNYILISKDVTGAT----KVKYYYA 1239
             :                |::                  :  :|::   ::
Db  2670 QTVTVTLNGATYGTVQADGSWSVSVPTSALGALNASNYTVSATVNDKAGNPGSASHNLA 2729
Qy  1240 VNDGFP---------GDH------YAVMISKT--GTNAGDFTVVFEETPNGINKGGARF 1281
             |:                        :   :::                       ||
Db  2730 VDTTAPVLTINTVAGDDIINDAEHAQALVISGTSSGGEAGDVVSVVI-----LNGKTY 2781
Qy  1282 GLSTEANGAKPQSVWIERTVDLPAG------------TKYVAFRHYNCSD--------- 1319
             :                                     |:||
Db  2782 TTTLDASGN-----W---SVGVPAADVTALGSGAQTITASVSDRAGNSDDASRTVTVSLS 2833
Qy  1320 -----LNYILLDDI---------QFTMGGSPTPTDYTYTVYRDG--------TKIKEGLTETT 1360
             :  ::|                    ::                  :   |:  |:
Db  2834 APVISINTIAGDDVINATEKGSDIALSGTSDQPAGTAITVTLNGQNYSATTDASGNWSVT 2893
Qy  1361 FEEDGV-ATGNHEYCVEVKYT--AGVSPKKCVDVTVNST----QFNPVQNLTABQAPNSM 1413
             |:|                        |||
Db  2894 VPASAVSALGEATYSVATSVTNAQGNSSTASHNVQVNTALPGIITINPV------AT 2943
Qy  1414 DAILKWNAPASKRAEVLNEDFENGIPASWKTIDADGD---GNMWTTTPPGGGSF 1465
             |::                                      :  :|:
Db  2944 DDII---NASEAGSAQTISGQVTGAAAGSTVTVELGGKTYVTATVQADLSWNVSVPAADWQA 3001
Qy  1466 AGHNSAICVSSASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEH---YA 1522
                                                        :::|:::      ||
Db  3002 LG-NGELTV-NASVTNAVGNTGSGTRDITIDASLPGLRVDT--VAGDDVVNIIEHAQAQV 3057
Qy  1523 VYASSTGNDASNFANALLEEVLTAKTVV------TAPEAIRGTRAQ 1562
             ::                                  :: |:
Db  3058 ITGSSSGFAAGTALTVVINNQTYAATVLANGSWSVGVPATDVSNWPAGTLNITVSGANSA 3117
Qy  1563 GTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDD--------VVITSGN 1603
             :  :|||                               :  |:
Db  3118 GTQTSITHPLTVDLTAVAISMNSITSDDAINAAEKGAALTLSGSTSGVEAGQTVTVTFGG 3177
Qy  1604 APSYTYTYRNNTQIASGVTETTYRDPDLAT---GFVTYGVKVKVYPNGESAITEATALNIT 1660
             ::                                :  |::
Db  3178 K----TYT---TTVAANGSWSTTVPAADLAALRDGDASAQVRVTNVANGNSATATHEYSVD 3230
Qy  1661 SLADVTAQKPYTLTVTVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVVTAQGGHYAVMVVV 1720
             :::                                  :::|::
Db  3231 SAAPTV----TINTIASDNIINASEAAAGVTVSGTSTAQTGQTLITVTLNGINYQTTVQT 3285
Qy  1721 DG 1722
             ||
Db  3286 DG 3287
```

RESULT 7
B85547
probable RTX family exoprotein [imported] - Escherichia coli
C;Species: Escherichia coli (strain O157:H7, substrain
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

RESULT 8
AF1450
probable peptidoglycan bound protein (LPXTG motif) lin0141 [imported] - Listeria innocua
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AF1450
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kuraphat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1450
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1993 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95374.1; PID:g16412560; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0141

Query Match   3.4%;   Score 313;   DB 2;   Length 1993;
Best Local Similarity 20.3%; Pred. No. 8.5e-09;
Matches 434; Conservative 208; Mismatches 702; Indels 796; Gaps 114;

```
Qy    1 MRKLLLIAASLLGVGLYAQSAKIKLDAFTTRTTCTNNSFKQFDASFSNEVELTKVETK 60
          :|: ||||:||||||:                                  :  |   :
Db    7 LRRPLLLVMAALLIIG---------QLNL-----------------------SSFRVF 36 AEEK

Qy   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQYSLNQYGSEKL 120
Db   37 GNESVSYEI-----QNELSSDKKKAKLKIKTTPKNE-----EVKILTIQ 75

Qy  121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
          :|:                           :|:
Db   76 ---TPDGKKVE---------------GQE------------------------AEYT 90

Qy  181 VVANQLKVRNNIEIEVSPQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
Db   91 ATKN------GTVDFVINYQDQDK--TEKTYTASYEVNDISEEVQNTPKPTVTEG----NT 139

Qy  241 PVRMLVVAGAKFKEALKPWLTWKA--------QKGFYLDVHYTDEAEVGTTNASIKAFIH 292
Db  140 PKK----GSNLLKSSAPTVSMKIPDYNKTNWENGDIKDVSVT--VEFNNNNAS-----G 187

Qy  293 KKYN----DGLAASAAPVFLALVG---DTDVISGEKGKKTKKVTDLYYSAVDGDYFPEM 344
          :|:|    ||:                          :||:|
Db  188 KKINFTLPDGMRFVSLPVPSSYQAPSSVDSNILS------------YFGA-GNPVGDS 232

Qy  345 YTFRMSASSPEELTNIIDKVLMYEKATMPDKSY----------LEKVLL-- 383
Db  233 IT----SVIVPNKETG--------YNKATYGTLSYELDPATEKLTINFSVQVDAAKYYGPTD 282

Qy  384 ------------------IAGADYSWNSQVGQPTIKYGMQYYNQ--EHGYT------ 415
Db  283 LKSPIKVDAYMGEGSTPAASAEQSIRAE-GKTVVGYANQKQVDTMFRNWYTSSFLPEVSP 341

Qy  416 ------DVVNYLKAPYT-----------GCYSHLNTGVSFANYTAHGSETAWA----DPLL 455
Db  342 STATEDSYNYTK-PYSVVNGISOMDGRGSKIFVPKNVKTTLYYPEGMEYVGVVNENKSLL 400

Qy  456 TTSQLKALT--NKDKYFLAIGNCCITAQFDYVQP-CFG--EVITRVK------EKGAY 502
          :|:|                                            :        ::
Db  401 TNNENRITTHYPENRK----------VEIDFKQENYYGIIETILAVKYKVPEGTEBGTY 449

Qy  503 AYIGSSP----------NSYWGEDYWSVGA-----------NAVFGVQPTFEGTSMGSYDATFLED 548
Db  450 TSPKVPHAVITTY--DDKVFETDALTNDASDTTTLAAKDTCKVVGKAANKMVMLPR 503
```

```
Db 2318 AAGNTASSQANLLVDSGLPDVTINT--------------------EAGADQTIS 2361
Qy  866 --MRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEV 923
Db 2362 GVVTRAAAGDTVTVTLGGN---TYTATVQ--SNLSWSVSVPTADLQALGNGDLTITASV 2415
Qy  924 KYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSV-----GQKVTLKWDAPNGTPNPNPNP 977
Db 2416 TNANGNTGSGTRDITIDANLPGLRVDTVAGDDIVNSIEHGQALVI----------TGGSS 2465
Qy  978 NPNPGTTLSESFENGIPASWKTIDADGDHGWKPG-NAPGIAGYNSNGC---VYSESFGL 1033
Qy 1034 GGIGVLTPDNYLITPALDLPNGGKLTFWCAQDANYASE--HYAVYASSTGNDASNFTN 1090
Db 2466 GLNAGAVLTVTI-NSVAYS-ATVQADG---SWSVGIPAANVSAWPAGPLTVEVDGQSSAN 2520
Qy 1091 ALLEETTTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1150
Db 2521 NPVSVSHPFTVDLTAVAISIN------TVASDDVINAAEKGTNLTLSGSTSGIEBSGQTV- 2573
Db 2574 ----TVIFGGKTYTASVAA--NGSW----SVNVPA-----------ADLATLPEGAANV 2611
Qy 1151 KANGKRADFTETEBSSTHG---EAPAEWTTIDADGDG-------------QWLCLSSG 1193
Db 2612 QASVSSASGNSA--SATHAYSVDASAPTLTNTIASDIILNAAEAGSPLTISGTSTAETG 2669
Qy 1194 QLDKLTAHGGS---NVVSSFSWNGM-----------ALNPDNVLISKDVTGATKVKYYAVNDG 1243
Db 2670 QTVTVTLNGATYTGTVQADGSWSVSVPTSALGALNASNYTVSATVNDK-----------AGNPG 2722
Qy 1244 FPGDHYAV---MISKTGTNAGDFTV--------------VFEETPNGINKG-------GARFG 1282
Db 2723 SASHNLAVDTAPVLTINTVAGDDIINDAEHAQALVISCTSSSGGEXGDVVSVVLNGKTYT 2782
Qy 1283 LSTEANGAKPQSVWIERTVDLPAG-------------TKYVAFRHYNCSD----------- 1319
Db 2783 TTLDASGN------W---SVGVPAADVTALGSGACTITASVSDRAGNSDDASRTVTVSLSA 2834
Qy 1320 ---LNYILLDDI--------QFTMGGSPTPTDYTYTVYRDG-------TKIKEGLTETTP 1361
Db 2835 PVISINTIAGDDVINATEKGSDLALSGTSDQPAGTAITVTLNGQNYSATTDASGNWSVTV 2894
Qy 1362 EEDGV-ATGNHEYCVEVKYT---AGVSPKKCVDVTVNST----QFNPVQNLTAEQAPNSMD 1414
Db 2895 PASAVSALGEATYSVTASVTNAQGNSSTASHNVQVNTALPGITTINPV----------ATD 2944
Qy 1415 AILKWNAPASKRAEVLNEDFENGIPASWKTIDADGD-----------GNNWTTTPPPGGSSFA 1466
Db 2945 DII--NASEAGSAQTISGQVTGAAAGSTVTVELGGKTYTATVQADLSWNVSVPAADWQAL 3002
Qy 1467 GHNSAICVSSASHINFEGQNPDNYLVTPELSLPGGGTLIFWCAQDANYASEH---YAV 1523
Db 3003 G-NGELITV-NASVTNAVGNTGSGTRDITIDASLPGLRVDT-VAGDDVVNIIEHAQAQVI 3058
Qy 1524 YASSTGNDASNFANALLEEVLTAKTVV---------------TAPEAIRGTRAQG 1563
Db 3059 TGSSSGFAAGTALTVVINNQTYAATVLANGSWSVGVPATDVSNWPAGTLNIITVSGANSAG 3118
Qy 1564 TWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDD-----------------VVITSGNA 1604
Db 3119 TQTSITHPLTVDLTAVAISMNSITSDDAINAAEKGAALTLSGSTSGVEAGQTVTVTFGGK 3178
Qy 1605 PSYTYTYIRNNTQIASGVTETTYRDPDLAT------GFYTYGVKVVYPNGESAIETATLNITS 1661
Db 3179 ---TYT---TTVAANGSWSTTVPAADIAALRDGDASAQVRVTNVNGNSATATHEYSVDS 3231
Qy 1662 LADVTAQKPYLTIVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVD 1721
Db 3232 AAPTV-----TINTIASDNIINASEAAAGVTVSGTSTAQTGQTLTVLNGTNYQTTVQTD 3286
Qy 1722 G 1722
Db 3287 G 3287
```

```
Db   1444  EATGTVDTTKAGEYPVTYKYN---DTTKTVTITVKDDATEI              1488
                 ||||                    |:: ||
Qy   1418  ----KWNAPASKRAEVLNEDFENGIPASWKTIDADGDNNWTTTPPPGGSFAGHNSAIC  1473
                 |:|
Db   1489  YTGDTWSA----------------KDNFDSA-------------VA            1512

Qy   1474  VSSASHINFEGPQNPDNYLVTPELSLPGGG-----TLTFWVCAQDANYASEHYAVYASSTGN  1530
                |||                              ||:|
Db   1513  LSKVTVINTVNTAQGTYPIT----YTYGGVSKTITVTVKENKKGINAHNATIYVGDSWT     1568

Qy   1531  DASNFANALLEE----VLTAKTVVTAPEAIRGTRAQGTWY-------QKTVQLPAGTKVV    1579
                 :::                    ||
Db   1569  AEDNFDNAVDKDGNPVEFSKVTVTETPNVNTNKA-GTYQLKYTPDGASKTVTLITVKNIQT  1627

Qy   1580  AFRHFGCTDFF---W---INLDDVVITSGNAPSYTYTYRNNTQI-ASGVTETTYRDPPL     1632
                 :: |                      |:
Db   1628  AVNAHNSTVYVGETWEAKDNFDSARNKDGETVAFADVEVEGNVDMTVAGTYSITYK---    1683

Qy   1633  ATGFYTVGVKVVVPNGESAIE--------------TATLNITSLADVTAQKPYTLTVVGKTI  1680
                |||                              ||:|
Db   1684  YDGF-SKTIKTVVKNPQTAITAHDSVVYTGDNWSAKDNFDSAID-KAGKP----VAYRDI   1737

Qy   1681  TVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYAVMVVV   1720
                |                       |  |
Db   1738  TVE--EDPTVDLN----TPGTYSVTVYKYQGISKVVQITV  1770
```

```
Query Match             3.3%;  Score 300.5;  DB 2;  Length 2468;
Best Local Similarity   20.5%;  Pred. No. 5.9e-08;
Matches 305;  Conservative 140;  Mismatches 505;  Indels 535;  Gaps 74;
```

```
Qy   534  EGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGN-----------IGNITHIGAH      579
               :::                                      ||
Db    99  DGPLMAGY--IPQESLAGFESLTGAGVLGGMSAGTALLVGAAAIGAGVAISNSSGGGGG   155

Qy   580  YYWEAYHVLGDGSVMPYRAMPKTNTYTLPASLPQNQASYSIQASAGSYVAI--SKDGVLY   637
               ||                              |||
Db   156  --------------GGSSVPDTTPPKAAS---GLKIAPDGSSISGQAEAGASVGIDTNGDGKPD  203

Qy   638  GTGVANASGVATAVSMTKQITENGNYDVVIT--RSNVLPVIKQIQVGEPSPYQPVSNLTAT   695
               |||                              |||
Db   204  LTVIADANGNFTAPLNPPLLNGQTVTVVVTDPAGNASPP-AQVTAPDTTAPAPATDVQVA   262

Qy   696  TQGQKVTLKWEAPSAKKAEGSREVKRIGDG----LFVTIEPANDVRA-------------   744
               :::                              ||
Db   263  PDGSSVTGKAEPGSTVGVDTD-----------GDGQPDTTVVVGPGGSFEVPLNPPLTNGETVTV  316

Qy   745  LAADNVWGDNTGYQFLLDADHNTFGGSVIPATGPLFTGTASSNLYSANFEYLVPANADPVV   804
               |||                              |||
Db   317  IVTDPAGNNST--PVTVEAPDTT--APAPATDVQVAPDGSS-----------VTGNAEPGA   362
```

```
Qy   549  SY--NTVNSIMWAG-----NLAATHAGNIGNITHIGAHYWEAY--HVLGDGSV--------  593
              :::                                      ||
Db   504  NYYINPDNE-SWAGLIQINNRQTAGVKTNQIYQIKFDENWEAYTVNIPPFDGTYPGNKVKD  562

Qy   594  -------MPYR---AMPKTN--TYTLPAS--LPQNQASYSIQASAGSY-----------   627
                    |||||                    ||||
Db   563  VQYKTNLNSEYRTYDGTLPKTNGNKMLTLDATAVGLQEGEYFTEVKANVGDFSVGFTNID  622

Qy   628  ------VAISKDG---VLYGTGVANA-----SGVATVSMTKQITENGNYDV          664
              |||                           ||||
Db   623  TAAPFKAANSASYGIVKPGITSVQFDVNIWNADDEENTKVSGSSTYTVANNISTAANGTA  682

Qy   665  VITRSNYLPVIKQIQVGEPSPYQPVSNLTATTGGQ--------KVTLKWEA           707
              |||                                      ||:|
Db   683  SFYNSEGTP-IKTARAGE--------TVTTKASLVHDPYGTRSVLNNPEVYLRQLE      731

Qy   708  PSAKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGY---QFILDA    763
              |||                                      ||||
Db   732  GTTVKPSSIKLTDQDGKEVDFTVEAKTAKNGDKVYVIKTKDITVGEFVGYFSKKQYLNIS  791

Qy   764  DHNTF---------------GSVIPATGP-LF-----TGTASSNLYSANFE         793
              |||                                      ||||
Db   792  YNTTFDMTLSKSIHTDIQELLAWGGSNVISALGANVFLDNGLDVNQNGRDAERLLSTNTS  851

Qy   794  YLVPANADPVVTTQNIIVTGQG-----------EVVIPGGVYDYC--IIN----        830
              |||                                      ||||
Db   852  TLSVPKQDTVAVETFLNVAGEGIKAGVVEGDDSTVSYFTPGTDADYMVKINTSSSSASS   911

Qy   831  -----PEPASGKMWIAGDGG--NQPARYD-----DFTFEAGKKYTFTMRRAGMGDGTD   876
              |||                                      ||||
Db   912  LDIFIPIPKTGQ-----DFGTKFQSEPFKWDMKLNGEVPMTAQQKSQFDLTYT--ADATA  964

Qy   877  MEVEDDSPASYTYTV------------------TFEEDGVAAGNH-             917
              |||                                      ||||
Db   965  DNYTSDSIYSNTLTDYGKANMVRIKVKTKIESGESQTIKVPLKVDETFE--SATAGNKIG  1022

Qy   918  ------EYCVEVKYTAG-------------VSPKVCKDVTVEGSNEFA----PVQNL    951
              |||                                      ||||
Db  1023  ERDVYNPYYAVETNTYSGSIAGTRVGAELVIVEVAGKLFKDKNASGIYETAQGDTPLANE  1082

Qy   952  TGSSVGQKVTL-KWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGD------  1005
              |||                                      ||||
Db  1083  T-------VELYKWDEATSYEPFTKDGQNVTT-----------TTDADGKYKFDY     1120

Qy  1006  ---GHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFW  1061
              |||                                      ||||
Db  1121  NLGIGYGKYAVKFPEKAGYKHT------LQNVG--RDKDLNSAA--PNMGTBIGW      1165

Qy  1062  V------CAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTW  1116
              |||                                      ||||
Db  1166  VKDIDPAQPDAQHINAGYISYISYVPETD--------LKVNLNEKTVQEGKSLK-       1208

Qy  1117  RQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTRTFESSTHGEA-----  1171
              |||                                      ||||
Db  1209  -------------ITLPKV--------APTSGEAAEDTI              1226

Qy  1172  -PAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLIS-KDVT  1229
              |||                                      ||||
Db  1227  EPDIFQKIQANTDGYKWTTADASVATAKTLTDGSVAIVGGSTNGKTIRVTDLTITALKDIF  1286

Qy  1230  GATK----VKYYYAVNDG---------------FPGDHY-AVMISKTGTNAGDFTVFEETP  1271
              |||                                      ||||
Db  1287  GTEKKSTAPLYVTTANGKVAQKDELKIGATNFTLEYKDAVVLSEAQAVTKAKTAAFEEVK  1346

Qy  1272  NGIN----------------------KGGARPG----LSTEANGAKPQSVWIERTVDLPA  1305
              |||                                      ||||
Db  1347  NGVNSDAEDRTDVVKVDEQQLKALKNGSNRGGTYPLTFTLEKNGKE-----VETVIDVKV  1401

Qy  1306  GTKVVAFRHYNCSDLNYIILLDDIQFTMGGSPTPTYTVYRDGTKIKEGLTETTFBEDG    1365
              |||                                      ||||
Db  1402  E-------KDLTEVNAHDSTIYVGDNWRAADNF----DSALNKEG--ETLTFADI      1443

Qy  1366  VATGN------HEYCVEVKYTAGVSPKKCVDVTV--NSTQFNPVQNLTAEQAPNSMDAIL  1417
```

Qy   805 TTQNIIVTGGGE------VVIPGGVVDYCIITNPEPASGK---MWIAGDGGNQ-----PARYD 852
           ::: | :: :   :|| ::: |||| ::|: || :.|||
Db   363 TV-GVDTDGDGQPDTTVVVGPGGSFEVPL-NPPLTNGETVTIVITDPAGNSSTPVTAEAP 420

Qy   853 DF--------------------TFEAGKKYTFMRRAGMGDGTDM-EVEDDSPASYTYTVVRDG 895
           || : ::   :::|
Db   421 DFPDAPQVNASNGSVLSGTAEAGVTIVIT-----DGNGNPIGQTSADANGNWSFT----PG 472

Qy   896 TKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPV------Q 949
            | :| || | | | : | ::|
Db   473 SQLPDGTVVNVVARD--AAGNS-----------SP--ATSITVDGVAPNAPVVEPSNGS 516

Qy   950 NLTGSS-VGQKVTL---------------KWDAPNGTPNPN 974
                 :: : ||:| : | | |
Db   517 ELSGTAEPGSSVTLTDGNGNPIGQTTADANGNWSFTPSTPLPDGTVVNVVARDAAGNSSP 576

Qy   975 ----------PNPNPNGTTLSESFE----------NGIPASWKTIDADGDGHGW--- 1009
                      | || | |:|| ||
Db   577 PASVTVDAVAPATPTVDPSNGTTLSGTAEPGSSVTLTDGNGNPIG--QVTADGSGN-WTF 633

Qy   1010 ----------------------------KPGNAPGIAGYNSNGCVYS 1028
                                        :||| | ||| |
Db   634 TPSTPLPNGTVNATATDPSGNASSPASVTVDAVAPATPVVNPSNGTTLSGTABPGATVT 693

Qy   1029 ESFGLGG-IGVLTPD---NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGND 1084
            || :| :| ||| ::|
Db   694 LTDGNGNPIGQVTADGSGNWSFTPTTPLPNG--------------TVVNATATD 733

Qy   1085 ASNFTNALLEETI----------TAKGVRSPKAIRGRIQGTWR---QKTVDLPA 1125
           :||| | | ::| |:|:| |
Db   734 ASGNTSAGSSVIVDSVAPATPVINPSNGTTLSGTAEPGSSVTLTDGNGNPIGQVTAD--- 790

Qy   1126 GTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES-STHGEAPAEWTTIDADGDG 1184
           | ::| |:| |::| | | : ||
Db   791 GSGNWSFT--PSTPL----------ADGTVVNATATDPAGNTSGGQS---TTVDGVAPT 834

Qy   1185 QGWLCLSSGQLDWLTAHGGSNVV-------------SSFSW-----NGMALNPD 1220
                ::| |: | : ||
Db   835 TPTVNLSNGSSLSGTAEPGSTVILTDGNGNPIAEVTADGSGNWTYTPSTPIANGTVVN- 892

Qy   1221 NYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGAR 1280
                |::   :: : :: : ::|| |||
Db   893 --VVAQDAAGNSSPGASVTVDSVAPA---APVNPSNGT-------------TLSGTAEPGAT 937

Qy   1281 FGLSTEANG------AKPQSVWIERTVDLPAGTKVVAFRHYNCSDLNYILLDDIQFTMG 1333
            || :| | | | :| | |||| ::
Db   938 VTL--TDGNGNPIGQVTADGSGNW-----SFTPGTPLA----NGTVVNATASDPTGNTSA 986

Qy   1334 GSPTFPTDYTYTVYV-------RDGTKI----KEGLTETTFEEDG---------VATGNHEYCVE 1376
           | :| | | | || ||| :| | ||| |||
Db   987 PASTTVDSVAPAAPVVNPSNGAEISGTAEPGSSVTLTDGSGNPIGQVTADGSGNWSFTPS 1046

Qy   1377 VKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFEN 1436
            | || | |||| | || | |:| |||
Db   1047 TPLADG------TVVNATATDPAGN-TGGQGSTTVDAI---APATPTVNLSNGSSLS 1093

Qy   1437 GIPASWKT-IDADGGN------NWTTTP------------------EGDYLYLVFADNLKN 1458
           | ||:: ||||||| || ||| |||
Db   1094 GTAEPGSSVTILTDGNGNPIAEVTADGSGNWTYTPSTPIANGTVVNVVAQDASGNSSPPAT 1153

Qy   1459 ------PP------GGSSPAGHNSAICVSSASHINFEGPQNPDNYLVTPELSL 1499
                 || ||||| : :| |||
Db   1154 VTVDSSAPPAPVVNPSNGVVISGTAEAGATVTLTDAGGNPIGQVTADGSGNWSFTPGTPL 1213

Qy   1500 PGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAI--- 1556
             | |:|||||| |:||||||
Db   1214 ANGTVIV-----ATADTPGNTGPQAA-----TTVDAVAPAPVIDP 1250

Qy   1557 -RGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTI---- 1611
            || :: | : :: || | : | ||
Db   1251 SNGTTISGT----AEAGAK------------VILTDGNGNPIGETTADGS 1284

Qy   1612 ----YRNNTQIASG-VTETYRDPDLATGFYTYGVKVVYPNGESAIETATALNIT------- 1660
                :| | || | | ||| || | | | : |||| |
Db   1285 GNWSFTPGTPLANGTVVNAVAQDPAGNTG------------- 1334

Qy   1661 -SLADVTAQKPYTLTVVGKITVTCQGEAMIYDMNGRRLAAGRNT 1704
            :| | | | | |||:| | || |: | |
Db   1335 GNLLNGTAEP--------------GSTVTLT------DGNGNPI--GQTT 1362

Db   1285 GNWSFTPGTPLANGTVVNAVAQDPAGNTG------------------PQGSTTVDAVAPNTPVVNPSN 1334

RESULT 10
S76412
hypothetical protein slr0408 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S76412
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocysti
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S76412
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-4199 <KAN>
A;Cross-references: EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BAA18541.1; PID:g16536;
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: Synechocystis hypothetical protein slr0408

Query Match   3.2%; Score 295.5; DB 2; Length 4199;
Best Local Similarity 18.1%; Pred. No. 2.6e-07;
Matches 438; Conservative 281; Mismatches 752; Indels 955; Gaps 108;

Qy   4 LILLIAASLLGVGLY--------------------AQSAKIKLDAPTRTT----CTNNSFKQFDASFSFNE 51
           ||| | :|:::| | ||| | :: | | :| |
Db   533 LLGLIGARIAGFPTWTQTFAASSSQNNTGLEANNTASVDGFLLTINLAQFLEPNLSLNP 592

Qy   52 VELT-KVETKGGTFASVSIPGAFPTGEVGSPEVPAVRKL-IAVPVGATPVVRVKSFTEQV 109
            || | :| | |: : | | | ||| : | :| | || : : ||
Db   593 TDFTVTVNANGQISTIPVFGVVVQTDSQSNNSEVILRLEQEIPY---TTVNGQPFSNII 649

Qy   110 --YSLNQYGSEKLMP--------------------HQPSMSKSD 138
             :|| ||| :|:: | |:|:|
Db   650 VKYSGNQLGNFTSSPVALQSPDTLIYTYNPTSGTGNDYTQNNQQITIAFNGPLDPDILPT 709

Qy   139 ---VYNAAAYARKGFVGGELTQVEMLGTMRGVRIAALT-INP----VQYDVVAN---QL 186
              | |||| :|| || | | | | | | |: | | || || |:
Db   710 ENQLLNWFPAVTNSSNQAIAITDVAVNQSRIVLTLASNTAINPGEIYTVQYNPDANSDQQL 769

Qy   187 KVRNNIEIEV------------------SFQGADEVATQRLYDASFSPVFETAYKQLFNRD 229
           | ::: ::|| || | | ::|: || :: | |:| ||
Db   770 KAANQKTITAFTVSNGTAPTAWGLINRTFAGVINVQTPVLSNLTSDFPAQDTSPAL----- 825

Qy   230 VVTDHGDLY----------------------NTPVRMLVVAGAKFKEALKPWLTWKAAQKGFYLDVHYTDEAEV 280
            ||| | | | ||:| | | | : | ||:||:| : ||| | | |||
Db   826 ALTSQGDILLAWSSDTPPITPISVLA---------------EGDYLYLVFADNLKN 866

Qy   281 GTTNASIKAFIHKKYNDGLAASAAPVFALVGDTDVISGEKGKTKKVTDLYYSAVDGDY 340
           | ::| : |: : ||: | || |: | : | | ::|| :|||
Db   867 DSANPPSNSQFTIKTSDG--NTTTPINVSLAQNTITILTLNSVNASQIVEVSYSLSGTNL 924

Qy   341 FPEMY----------------------TFRMSASSPEELTNIIDKVLMYEKATMPDKSYLE----- 379
           | |: | :|:|| | ||::|:|| :| | |: ||
Db   925 TSNLYLADATNTSFWVPDFTNSVQSAGSTSTAPTSLLGSVIS-NLITIPFNQTLNVVNQIP 983

Qy   380 ----------------KVLLIAGADYS---WNSQVGQ---PTIKYGMQYYYNQEHGYT 415
                           ||::|||| | | | | | | | ||:|| |:|
Db   984 NGGQFQVTVNGNSTSPITVTSVTVADTSITLVLNQIIGQGQLVTIT-----YTPNSNGEN 1038

Qy   416 DVYNYLKAPYTGCYSHLNTGVSFANYTAHGSETAW 454
            : : | || : :| | |
Db   1039 NLVSNTGTPQTVASFSTNNFLTTASSTGTVIKTAFSPFGDSGISSITTIPGTGINSDVV 1098

```
Db  2079 -DLDGDGHQDLAISA------LGVNGSAGAV--YALSGGSKFTPSSLQALNEAG-----N 2124
Qy  1238 YAVNDGFPGDHYAVMISKTGTNAG-----DFTVVFEETPNGINKGGARFG---LSTE--- 1286
Db  2125 LIINGGIAGGRAGMTIMTPGDVNGDGYCDFLITAPQAGNGTGQSYLLFGPLDLSTEIVPI 2184
Qy  1287 ----ANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD 1340
Db  2185 IELNAIANDSKQVFVLNGSLPNQLAGTAVVS-------LGNITGTQGPNNRPID 2231
Qy  1341 -----YIYTVYR-------DGTKIKEGLTETTFEEDG--- 1365
Db  2232 SFLISAPNAQQFYVVFGQPWLAADGSLNLADVASDNGFVIDGNLI--GNPPTTFETTSQY 2289
Qy  1366 ------VATGNHEYCV------EVKYTAGVSPKKCVDVTVNSTQ------F 1398
Db  2290 IDTTPAILINGSNLYLAYKGFGGNNQIYFTVSTNNGQSWNSEVQLPQSAQTIFPPAIAFF 2349
Qy  1399 NPVQNLTAEQAPNSMDAILK-------WNAP------ 1422
Db  2350 NNVLYLAVVDGNNGLNIITSQDQGQTWNAPLALGGTSSTPPTLFVYQGTLSLLPAANNST 2409
Qy  1423 ------ASKRAEVLNED---FENGIPASWKTID--- 1446
Db  2410 STVLQFYLNSSNEWIYANEIGSNQTAISAISATVLGDTLYLVYKGGTRNTPSTLDYITST 2469
Qy  1447 --ADGGNNWTTTPPPGGSSFAG------HNSAICVSSASHINFEGPQ- 1486
Db  2470 TNADLSANDWSSIPIPGVSSQGGPSLTNDGTNLYLSYLDSSNQLNFVSSGNGINWSSPQV 2529
Qy  1487 ------NPDNYLVTP------EL-------SLPGGGTLTFWVCAQDA 1514
Db  2530 ITNNISQSPPAIAFANNELYLSYPGPQGSQELNVTSFFLPFTGSILGNGSLVRFL--GDV 2587
Qy  1515 NYASEHYA-VYASSTGNDASNFANALLEVLTAK---TVVTAPEA-IRGTRAQGTWYQK 1568
Db  2588 N--GDGFADVFSGGTNAGAIIFGNSTKDLLTTASGSEDLVISVPNATLRDVISVG---- 2640
Qy  1569 TVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYR 1628
Db  2641 -----DFNGDGIKDLGVLDGNGNFY---VVLGNTSLG----------- 2669
Qy  1629 DPDLATGFYTYGVKVVYPNGESAIETATINITSLADVTAQKPYTLTVVGKTITVTCQG-- 1686
Db  2670 -----DLKTILSITS------SSSPVVINQVGGVTKSMAIGDY 2700
Qy  1687 ------EAMIYDMNGRRLAAGRNTVV 1706
Db  2701 NGDGYDDVLLWGDNGNQVAWGNSTGV 2726
```

Query Match   3.1%;   Score 280.5;   DB 2;   Length 3029;

```
Qy   455 LTTSQ---LKAL------TNKDKYFLAIGNCCITAQFDYVQPCFGEVITRVKEKGAYAYI 505
Db  1099 ATLYQNPSTKALQNVVAWNVNVDTSAISLK----TIPGQNYGPNEAALITTAAQQSDIYYA 1154
Qy   506 GSSPNSYWGEDYYWSVGANAVFGVQP--------- 531
Db  1155 VLGPDNQWGL----AAPIFSSQPGQDQKVTLGVGPGGNLLAAWLNTQLDSDGDPNTT 1207
Qy   532 ---TFEGTS------MGSYDA----TFLEDSYNTVN---SIMWA--------- 559
Db  1208 IQLATFNGTTWTNPTILGGANAGINPNSFSELSISSINGQPAIFWTESRPPSYSNLVSEQ 1267
Qy   560 ------GNLAATHAGNIGNITHIGAHYVWEAYHVLGDGSVMPYRAMPKTNT------ 604
Db  1268 NPLVYLRLGELSGTTVINQGQLSVAGNGTYSTAGYTIGQ--VGALENTNNTGDFNP 1322
Qy   605 ------YTLPASLPQNQASYSIQASAGSYYAISKDGVLYGTGVANASGVATVSMTKQ- 655
Db  1323 AVLFNGGGITINSPVPVSVQGFSVEF---WFKLPTSDGVV---GLANLAGVFDLSLNEDS 1376
Qy   656 ---ITENGNYDVV----ITRSNYLPVIKQIQVGEPSPYQPV--------- 689
Db  1377 LTFTLNNGSNPQISGTVTTGNWHYVV------GTYDPVKQILDLYLDGQLVNTLENIA 1428
Qy   690 ------SNLTATTQGQKVTL------KWEAPASAK----KAEGSREVKRIG 723
Db  1429 FANLPQSGTLTLAGSGGSVYLDEFAFYNSILSYVDNGSSPSSSNNNFNLNLTGSQLI--- 1484
Qy   724 DGLFVTIEPANDVRANEAKVVLAAD---NVWGDNTGYQFLLDADHNTFGSVIP------A 774
Db  1485 NGIWGVNEVGSHYQARFEPVTAGPETNYSVN-DSSGNSWQSPVSINPVDEVVPTILSAA 1543
Qy   775 TGPLFTGTASSNLYSANFEYLVPANADPVVTQNIIVTGQGEVIPGGVYDYCITNPEPA 834
Db  1544 NNPIW-DIVSANPAGNNNAQIAP-NGNPDTIFQ-VNLTGQQSEITG---FTVTT--S 1593
Qy   835 SGKMWIAGDGG--------NQPARYDDFTFEAGKKY------TFTMRRA 869
Db  1594 NNQLWTVGTDGTGNAFSESWQLGVILAENADSTTPQLEFISGDKLLNSLNPGATFSHRVM 1653
Qy   870 GMGDGTDMEVEDD------SPASYTYTVYRDG------ 895
Db  1654 GATETFTIFVDTBGSPLTSPA--TVNILQGQTDPITFTSLSPIPNQGGPVSANSPDYLD 1711
Qy   896 ------TKIKEGLTATTFEEDG------VAAGNHEY---- 919
Db  1712 NQVLGIATIKEANDASLSLVDSGFVIDTDNPAIAAVMASGFSNGALAYVAVGNRGYTTQG 1771
Qy   920 ------CVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPN 974
Db  1772 NAVQGSIQILFAGG-------DVLSQKSTLPLTTTDLSGNGDGVLITGITDA--GDINNN 1822
Qy   975 ------PNPNPNGT------TLSESFENGI----PASWKT 999
Db  1823 VPMALVTGDVDGDGVDDLVIGNANANGGTGSIYYINGHYLNGLRGKNQIIDLSNASNWTS 1882
Qy  1000 ------IDADGDGHGWK-------PGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYL 1045
Db  1883 DQGFVIDGVDAEG-GAGFSVAIGNFTGNDPQIA-------FGAPFAAKNGNGVAVGKVYL 1933
Qy  1046 ITPA--------LDLPN------GGKLTFWVCAQDA 1067
Db  1934 VSPSNPSQLSPIIHIGNTFNLTNPQNPAQTVTVGETAGYSLGVSRKISGGPVTF-----T 1987
Qy  1068 NYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGT 1127
Db  1988 NNSGDDLFIGSSTYVQVSN------QWV-GKSALPSNNQGNYPDTTM-IAAGA 2033
Qy  1128 KYIVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAP------AEWTT 1177
Db  2034 VHV------YSQTSSQPFGKVATYTGPNIPAANGVGANYLAGAAISLGDFT--2078
Qy  1178 IDADGDGGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYY 1237
```

```
Best Local Similarity  18.9%;  Pred. No. 1.1e-06;
Matches 359;  Conservative 219;  Mismatches 705;  Indels 613;  Gaps 85;
```

Right column alignment:

```
Db 1916 VFNGTAEA------DSTVVVSLDGIEIGTVTANGAGEWTLDYTGTLLADGDYELSVTATN 1969
Qy  977 PNPNPGTTLSESFENGIPASWKTIDA---DGDGHGWKPGNAPGIAGYNSNGCVYSESFGL 1033
Db 1970 PTGN-SATATQTIVDTTAPTVTINA.AVDDIINAVEAGSPVAVSGTTTG------ 2018
Qy 1034 GGIGVLTPDNYLITPALD-----LPNGGKLTFWVCAQD-ANY--ASEHYAVYASSTGND 1084
Db 2019 -----VEDGQVVTVTIDGNTYTATVTGNAWTFNIPVADIANFEATEEVVATVSDLAGNP 2072
Qy 1085 AS-----NFT----NAL-LEEITTAKGVRSPKAIRGRIQGTNRQKTVDLPA-GTK 1128
Db 2073 ATPATRNITVDTTAPTVTINAIAVDDIINAVEAGSPVAVSGTTTGVEDGQVVTVTIDGNT 2132
Qy 1129 YVA------FRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGRAPA 1173
Db 2133 YTATVTGNAWTFNIPVADIANFEATEEVVATVSDL--AGNPATPATRNITVDT--VAPA 2187
Qy 1174 EWTTIDADCDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNVLISKDVTGATK 1233
Db 2188 --VTIDSISDDTG-----AQANDFITND-----DTLVFNGTAEADSTVVVSLD------ 2228
Qy 1234 VKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQ 1293
Db 2229 -----GIEIG-------TVTANGAGEWTLDY-- 2269
Qy 1294 SVWIERT--VDLPAGTKVVAFRHYNCSDLNYILLDD-IQFTMCGGSPTPDTYTVYYRDGT 1350
Db 2270 SATATQTIVVDTTAPT------VTINAIAVDDIINAVEAGSPVAVSGTTTGVEDGQ 2319
Qy 1351 KIKEGLTETTFEEDGVATGN-HEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQA 1409
Db 2320 VVTVTIDGNTY--TATVTGNAWTFNIPVADIANFEATEEVVATVSDLAGNPA-------T 2370
Qy 1410 PNSMDAILKWNAPASKRAEVLNEDFENGIPA----------SWKTIDADGDGNWT 1455
Db 2371 PATRNITVDTTAPTVTINAIAVDDIINAVEAGSPVAVSGTTTGVEDGQVVTVTIDGNTYT 2430
Qy 1456 TTPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDAN 1515
Db 2431 AT-------VTGNAWTFN;PVADIANFEATE------ 2454
Qy 1516 YASEHYAVYASSTGNDASNFA-NALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPA 1574
Db 2455 --EVVATVSDLAGNPATPATRNITVDTVAPAVTIDSISD--DTGAQ 2496
Qy 1575 GTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTI---YRNNTQIASGVTETT--Y 1627
Db 2497 -----ANDF--ITNDDTLVFNGTAEADSTVVVSLDGIEIGTVTANGAGEWTLDY 2543
Qy 1628 RDPDLATGFYTYGVKVVYPNGESAIETATLNITSLAD------VTAQKP-- 1670
Db 2544 TGTLLADGDYELSVTATNPTGNSATATQTIVDTTAPTVTINAIAVDDIINAVEAGSPVVA 2603
Qy 1671 ------YTLTVVGKTIIVTCQGEAMIYDM 1693
Db 2604 VSGTTTGVEDGQVVTVTIDGNTYTATVTGNAWTFNI 2639
```

Left column alignment:

```
Qy   20 QSAKIKLDAPTRTTCTNNSFKQFDASFSFNEVELTKVETKGGTFASVSIPGAFPTGEVG 79
Db 1135 QVVTVTIDGNTYTATVTGNAW-----TFNIPVADIANFEATEEVVATVS------DLAG 1182
Qy   80 SPEVPAVRKLIAVPVGAT-PVVRVKSFTEQVYSLNQYGSEKLMPHQPSMSKSDDPEKVPF 138
Db 1183 NPATPATRN---ITVDTTAPTVTINAIA------VDD 1210
Qy  139 VYNAAAYARKGFVGQELTQVEMLGTMRGV----RIAALTINPVQYDVVANQLKVRNNIEIE 195
Db 1211 IINAV------EAGSPVAVSGTTTGVEDGQVVTVTIDGNTYTATVTGNAWTFNIPVA 1261
Qy  196 --VSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVVTDHGDLYNTP----VRMLVVAG 249
Db 1262 DIANFEATEEVV----ATVS------DLAGNPATPATRNITV- 1293
Qy  250 AKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYND------GLAASA 303
Db 1294 ----DTVAPAVT-------IDSISDD-----TGAQANDFI---TNDDTLVFNGTAEAD 1332
Qy  304 APVFLALVGDT--DVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNII 361
Db 1333 STVVVSLDGIEIGTVTANGAGEWTLDYTGTLLA--DGD-----YELSVTATNPTGNSATA 1385
Qy  362 DKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGTDVVNYL 421
Db 1386 ------PTVTI------NAIAVDDIINAV 1412
Qy  422 KAPYTGCYSHLNTGVSFA---------NYTAHGSETAWADPLLTTSQLKALTNKDKYFL 471
Db 1413 EAGSPVAVSGTTTGVEDGQVVTVTIDGNTYTATVTGNAWT------FNI 1455
Qy  472 AIGNCCITAQFDVQOPCFGEVITRVKEKGAYAYIGSSP------NSY 512
Db 1456 PVAD---IANFEATE----EVVATVSD------LAGNPATPATRNITVDTTAPTVTINAI 1502
Qy  513 WGEDYYWSVGANAVFGVQOPTFEGTSMGSYDATFLE-DSYN-TVNSIMWAGNLAATHAGNI 570
Db 1503 AVDDIINAVEAGSPVAVSGTTTGVEDGQVVTVTIDGNTYTATVTGNAWTFNIPVADIANF 1562
Qy  571 GNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNTYTLPASLP------- 612
Db 1563 EATEEVVA------TVSDLAGNP--ATPATRNITVDTVAPAVTIDSISDDTGAQANDF 1612
Qy  613 ---QNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRS 669
Db 1613 ITNDDTLVFNGTAEADSTVVVSLDGIEIGTVTANGAGEWTLDYTGTLLADGDYELSVTAT 1672
Qy  670 NYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVKRIGDGLFVT 729
Db 1673 N------PTGN-SATATQTIVDTTAPT------VT 1695
Qy  730 IEPA--NDVRANEAKVVLAADNVWGDNTGYQFILDADHNTFGSVIPAT--GPLFTGTAS 784
Db 1696 INAIAVDDIINAVEAGSPVA--VSGTTTGVE------DGQVVTVTIDGNTYTATVTVT 1743
Qy  785 SNLYS------ANFEYLVPANADPVVT------TQNIIVTGQGEVVIPGGVY 824
Db 1744 GNAWTFNIPVADIANFE---ATEEVVATVSDLAGNPATPATRNITVDTTAPTVTINAIA 1799
Qy  825 DYCITNFPEPASGKMWIAG------DGGNQPARYDDFTEA--GKKYTFMRRAGMGD-- 873
Db 1800 VDDIINAVEAGSPVAVSGTTTGVEDGQVVTVTIDGNTYTATVTGNAWTFNIPVADIANFE 1859
Qy  874 ------GTDMEVEDDSPASYITYTVYRDGTKIKEGLTATTFEEDGVAAGN---HEYCV 921
Db 1860 ATEEVVATVSDLAGNPATPATRNITV----DTVAPAVTIDSISDDTGAQANDFITNDDTL 1915
Qy  922 EVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV-----TLKWDAPNGTPNPNPN 976
```

```
A;Accession: AB1180
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2044 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98920.1; PID:g16410230; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0842

Query Match          3.0%;   Score 276.5;  DB 2;  Length 2044;
Best Local Similarity  19.8%;   Pred. No. 9.4e-07;
Matches  368;  Conservative 211;  Mismatches 688;  Indels 591;  Gaps 89;

Qy  272 VHYTDEAEVGTTNASIKAFIHKKYN----DGLAASAAPV-------FLALVGDT 314
Db  912 VAAGNH--EYCVEVKYTAGVSPKVCKDVTVEGSNEPAPVQNLTGSSVGQKVTLKWDA-- 966

Qy  194 VNFGDSTSTG----------KKVNFTLPDGMRFVSLPVSNYQAGTNEDTGVLSYLGAS 242
Db 1029 ATEGNKIGERDIYNPYYRVI-------TNTFA--GSLSGTKVGAELVIGEVAGML 1074

Qy  315 D-----VISGEKGKKTKKVTDLYISAVDGDYFP--EMYTFRMSAS---------SPPELTNI 360
Db  967 -----PNGTPNPNPNPNPFGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYN 1021

Qy  243 DPLGIAITSVEVPNKETAYSQATFGTVSYELSPGTEKASFKFSVRVDAAKYYGPTDLADP 302
Db 1075 FNDKDVNGSYEKDKGDEPLANETV-ELYKWNVSTS----------EYEPAKVGDKNITATTD 1125

Qy  361 IDKVLMYEKATMPDKSYLEKVLLIAGADYS----WNSQVG---QPTIKYGMQYYNQEH- 412
Db 1022 SNGCV--YESESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYA 1078

Qy  303 IKTEVYKGEASTP----------VASADQAIRAEGNKVVGYADQNHVKIMFRTWYNDQRL 352
Db 1126 SNGKYSFDYSDGVGYGNYAVKFPDK---------AGYQYTL- 1157

Qy  413 ------GYTDVYNLKAPYTGCYSHLNTGVSF-----ANYTAHGSETAWADPLLTTSQL 460
Db 1079 SSTGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKY--VAFRHFQ 1136

Qy  353 SEVLASTDTTDSYNYTKS------YSVVNGSNSLDSRGAAVVIAKNIEVTLYYP-EGMEFV 406
Db 1158 KNIGKDTS-----------LNSDVPYSGADRGWAKQINPTLPS-SQYTNAGYYAYN 1201

Qy  461 KALTNKDKYFLAIGNCCITAQFDYYQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWS 520
Db 1137 STDMFYIDLDEVEIKANGKRADFTETFESSTHGEA-----------PABWTTIDADGDGQWLCL 1190

Qy  407 NVVNNAGTVLKENSNVTIT---NY--PSENKVVINNKHLNNSA----TSNSIYGVKYKVP 457
Db 1202 PTQDLKVNLDEBKQVQM-GRSLBEITLPKVASTTGQAAEDTIEPSFFKNIKATTNGYKWTVA 1260

Qy  521 VG---------ANAVFGVQPTFEGTSMGSYDATFLEDSYNTV- 553
Db 1191 SSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLIS-KDVTG--ATKVKYYAVNDGFPG 1246

Qy  458 KGTPAGTYSTAKAPHAVITTYDGEVFESDALTTNPNDLITIIAPLDTCKVVDTTENKMTLT 517
Db 1261 DTKVATVQTLADGSAAVVGVSTNNKTIDVTDLTITIQDIFGTEKSSKAPVVTGTDGTVA 1320

Qy  554 ------NSIMWAGNLAATHAGNIG-----NITHIGAHYWEAYHV- 587
Db 1247 --DHYAVMISKTGTNAGDF-----------TVVFEETPNGINKGGA- 1279

Qy  518 TANGQLNPNNETWAGSIQINNKKSAGVKKNOMYQIKFDPNWEAYMVNIPFDSTISGNKIS 577
Db 1321 QQDGYTM-----------GATDFSLEYKEATDLTKAQALTLAKTAAFEEVKDGVNSSAEDRLDQ 1373

Qy  588 ----LGDGSVMPYRAMPKTNT--YTLPA--SLPQNQASYSIQASAGSY----------- 627
Db 1280 -----RFGLSTEANGAKPQSVW-IERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGG 1334

Qy  578 EVQYKTNLNDAFRTFDGALIKNNNQMYRLDAKAVGLEEGEYFTEVKANVGFDFAPGYQSTE 637
Db 1374 VQVNQTQLDAIKNGSNQGGVYPLTYTITKDSKITVSVTIQVTVAKDLTAVNAHDSTIYLGD 1433

Qy  628 ------VAISKDGVLYGT--GVANASGVAT--VSMTKQITENG- 660
Db 1335 SPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCV-----EVKYT-AGVSPKKCV 1389

Qy  638 ASATYRWNSTASYGKIKPGVTSVQYEGAIWDADDESNTKVSGISTYKVSSAESTVANGTA 697
Db 1434 TWDAEDNF--------DSAANKEGDTNVAF-SDVTVTGSVNTAVAGSYPVTYTYNGVSKK--I 1485

Qy  661 ------NYDVVITRSNYLPVIKQIQVGEPSPY----QPVSNL 692
Db 1390 DVTVNSTQOFNPVQNLITAEQAPNS-----------MDAILKWNAPASKRAEVLNEDFEN 1436

Qy  698 AFYNKAGTKVKTASAGETINTKATLVIHDY-PYGTRTVLNDPEVYLHALEGTTIQP-SSI 755
Db 1486 NVTVK----------DKLTAVNAHDSTIYTGDTWNAGDNFDSALDKDGNSVAFADITVTGTVN 1538

Qy  693 TATTQSQK-VTLKWEAPSAKKAEGSREVKR--IGDGLFV----------TIEPANDVRA 738
Db 1437 GIPASWKTI-DADGDGNNWTTTPPPGGGSSFAGHNSAICVSSA--SHINFEGFQONPDNYL 1492

Qy  756 KLTDQGGKDVNFTVEQETANNGDKVYALKTTDVSVGVYGYPTKHKYLNISYDTTFDVTL 815
Db 1539 TNQAGTNTITYSYDGVSKTITVTVLENKEGISAHDSTIYGDAWNAKDNFDSAFDKDGNA 1598

Qy  739 N-----EAKVLAADNWGDNTGYQFLLDADHNTFGSVIPATGP-LFT----------GTA 783
Db 1493 V-----------TPELSLPGGGT----------LTFWVCAQDANYASEHYAVYASSTGN 1530

Qy  816 NKSINMDAQQVIA---------WGAN--------VTPATGTNTFSDLGLDVNKNGRD 855
Db 1599 VDLEDVTVEKFTVDTTKAGAYEVTYKYGKVSKKITITLTVKAKLTAVNAHDSAIYIGDTWS 1658

Qy  784 SSNLYSANFEYLLVPANADPVVTTQNIITVTGQGEVVI----------PGGVYDYC--I 828
Db 1531 DASNFANALLEEVLTAKTVVTAPEAIRGTRAQCTWYQKTVQLPAGTKYVAFRHFGCTDFF 1590

Qy  856 NDKLLSVNSSTISISIPKQDTVTVETFLSVAGEGDKAAYMEGEDSTVSYFSPGTDADYTVKI 915
Db 1659 AEDNFDSALDKD---GNSVAFADIEVKGTVDTD-------------KVGTYPVTYTYDGVSKTI 1706

Qy  829 TN-----------PEPASGKMWIAGDGG----NQPARYD----------DFTFEAGKKYTF 864
Db 1591 WINLDDVVITSGNAPSYTYTIY----------RNNTQIASGVTETTYRDPDL----ATGFY 1637

Qy  916 TNTSSGNASTFELYIPIPKTGQ----DFGSKFQSEPFKWDMKLNGALPVSAEQQDQFDV 970
Db 1707 NIQVKD-ILTAVNA--HDSEIYIGDNWNAKDNFDSAKDKDGNTVNWKDINVSENPAVDLE 1763

Qy  865 TMRRAGMGDGTD-MEVEDDSPASY---TYTVVRDGTKIKEGLTAT-------TFEEDG 911
Db 1638 TYGV-KVVYPNGESAIFTATLNITSLADVTAQKPYTLT----------GTA 1674

Qy  971 SYATTATGDNYDSTDIYSNTVADYEKVNMVRIKVKTQINAGETQTFKVPLKVDETF-DS 1028
Db 1764 TVGVYQVTYSYGG--VSKTINLTVHPRKTSLEVHDSTMYTGDKWAEDNFDNATDKKGD 1820

Db 1675 -VVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAV 1731

Db 1821 QIPFKDVTTGQ---------------VDSKTAGTYEITYITYDGLKKVARITV 1858
```

RESULT 13

AF1489

cell wall-associated protein precursor wapA (B. subtilis) homolog lin0454 [imported] - I

C;Species: Listeria innocua

C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C;Accession: AF1489

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, H.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

```
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2167 <GLA>
A;Cross-references: GB:AU592022; PIDN:CAC95686.1; PID:gi6412895; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0454

Query Match          2.9%;    Score 269;   DB 2;   Length 2167;
Best Local Similarity  19.2%;   Pred. No. 2.7e-06;
Matches 407;   Conservative 251;   Mismatches 741;   Indels 720;   Gaps 114;

Db   751 VRGGEVNATNGNFLFHETDFNLEGRGPSINVNRTFNSQDDATGIFGKGWTSTLEEKLVEE 810
Qy   660 GNYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPS----------AKKA 713
Db   811 ENGNIVWVESD---KKIH----------RFTKKGD---KYEAPPGIYSETTKNA 848
Qy   714 EG------SREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHN 766
Db   849 DGYLKIEEDKSETRFLVDG---RLKSEKDTKGNE----LTYEYTDGKLTS--LRDASGR 898
Qy   767 TF------GSVI-PATGP-------------LFTGTASSNLYSANFEYLVPANADPVVT 805
Db   899 TVTLTYEGELVKELVGPEDRKISYTYNDKQELISSSTARGKLYRYGY-------TDGLLT 951
Qy   806 TQNIIVTGQGEVVIPGGVDYDYCITNPEPASGKMWIAGDGGNQPARYDFTFEAGKKYTFT 865
Db   952 S-----------------IYDPKHTEEKPYETTF-------AYEEKLTEITDPVGKKTTLS 989
Qy   866 MRRAGMGDGTDMEVEDDSPASVTYT--------------VYRDGTKIKEGLTATTFBEDGVAAGN 916
Db   990 YDKA--EQQTTLTNEKSKKTIYSYNDAGNPKKEIVDADGLKL---TTTYTYESNNLVKEV 1044
Qy   917 HEYCVEVKYTAGVSPKVCKDVTVEGSNRFAPVQN--LTGS--SVGQKVTLKWDAPN--- 968
Db  1045 NPKGQEETYAYDADGNITKATDAYGTESYTYNDNDVTSSTDEGRKTTVTTDGADAVSE 1104
Qy   969 ----------GTPNPNPNPNPGTTLSES-FENGIPAS-WKTIDADGDHGWM 1009
Db  1105 TLATESQVSSVTQYDAYGNPIRGSGELSSGGNLLQNSGFEKGAGVSNWTLIQSDAKGSMT 1164
Qy  1010 --KPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDA 1067
Db  1165 FDNTQSAPG-----------ALGGGSGV------------KLTSBANSTVK 1192
Qy  1068 NYAS----------EHYAVIASSTGNDASNFTN--ALL----------BETTTAKGV-RSPKAIR 1109
Db  1193 GYSSVTQRVDVEPETTYTFSAWIKTSGMTNADALLIGRLQDANAKDVTDAGVWQSNRATS 1252
Qy  1110 GRIQGTW--RQKTVDLPAGTKYVAFRHFQSTDMFYID----------LDEVEIKAN 1153
Db  1253 IKKNGDWVKRQLIFPKTSKNTRQV----------LLYLDNEQPAPHKGKGTIWYDNVQFEKG 1303
Qy  1154 GKRADFTETFESS---THGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSF 1210
Db  1304 SVASSYNPVVNNSFENHNGTLPTGWMR--------TGNTALTQAKVVDNQSHSGDSAV-- 1353
Qy  1211 SWNGMALNPDNYLLSKDVTGATKVKYYYAVNDGFP--------GDHYAVMIS-------- K 1255
Db  1354 --------YFERK----ATSEAYTHIVQD-VPVNQKEAKALTISALSKSEDAKANGS 1397
Qy  1256 TGTNAGDF---TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVA 1311
Db  1398 VAIMSNDYSVWGTVYYQDGTTSSVQG--QFPLGT--NDWNRSAVVVKPTKPVKMIKVYTM 1453
Qy  1312 FRHYNCSDLNYILLDDIQFTMGGSPTPTDY------------TYTVYRDGTK 1351
Db  1454 FRN----GLTGKAWFDDVRVIEGEVLTKNEYDASGNYVTASYDEEGRKISFTYDIY--GNK 1508
Qy  1352 IKE----GLTET-TFBEDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTA 1406
Db  1509 TSETDEKGNKKTLTYDADNA--------LIDTKLANGTSVAYKIDDNGNTTE--KNVTA 1557
Qy  1407 EQAPNSMDAILKWNAPASKRAEVLNEDFENGIPASW---KTI--DADGDGNNWTTTPPP 1460
Db  1558 S--------GKTQKNIYEYDVDNKITAFTDALNRTIKYEYDAAGNETKAIMPN 1602
Qy  1461 GGSSFPAGHNSA----ICVSSASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDAN 1515
Db  1603 GRVTESTYDSADRMDGIKWNDKLAFKFQYDPNGNQTKVTDEIN----SIVTDKT 1652
Qy  1516 YASEHYAVVASSTGNDASNFAN------------ALLEEVLTAKTVVTAPEAIRG 1558
Db  1653 YDDANRITKVAERGDVSYTYKDRPTKDNKGKTDKVGEVAINHGDYTAKTSYTNDLDRN 1712

Qy    69 IPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVK--------------SFTEQVYS-- 111
Db    24 LPGYAPLAEDSPVGVDAPDK---KAGEQAPTEVKEERTENEVVFDNHDGSFTKQIFADP 79
Qy   112 LN------------QYGSEKLMPHQPS-----MSKSDDPBKVPFVYNAAAYARKGF 150
Db    80 INMEVDGDMKRIDANVEKEADSDMIVPKQTPLELGFLEKMED----------GAYQKLTK 129
Qy   151 VGQELTQVEMLGTMRGVRIAALTINP-------VQYDVVANQLKVRN----NIBIEVSF 198
Db   130 AGAEVT-FRLKGARTGENBQAVTDQPATYKENEVTYENVFPRTDLRHLTFPQSVKEDLVL 188
Qy   199 QGADEVAT------QRLYDASFSPYFETAYKQLFNRDVYT---------DHGDL 237
Db   189 QEPNQVDTYVYQIETKLDLELAENGDVLFKNKSDETMYTLPKPVMTDSNVGTETGEAAQS 248
Qy   238 YNTPVRMLVVAGAKFKEALK---PWLTWKAQK------GFYLDVHYT---DEAEVGTTN 284
Db   249 ENVSFKVKQLTKTVYELQLKVDTEWLNDDAREYPVYIDPSVRLDKITYNANINSAKPTETN 308
Qy   285 ASIK------AFIHK--KYNDGLAASAAPVFLALVGGTDVIS-GEKGKKTKKV------ 328
Db   309 IGSKLWDSGQNAYTLKLGKWDNSTGNNAA--FLKM--DTSTLNKATISKATILKVYNIWHM 364
Qy   329 -----TDLIYYSAVDGDYFPEMYIFR------MSASSPEELTNIIDKVLM--------- 366
Db   365 SPTVKNDLMYYEANAKWSPWQVTWNTVPPVTRLGSVNVGRGEWANLDVTKTVQAWASGSR 424
Qy   367 ---YEKATMPDKSYLEKVLLIAG----ADYSWNSQVGQPTIKYGMQYYNQBHGY 414
Db   425 VNNGFRLGTNINKNYWKKVVASENNKNYPYLEVNYTY-AQPEKPTVKITSSN-GVGTGTGY 482
Qy   415 TDVYNYLKAPYTGCYS-HLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAI 473
Db   483 MDL-SWKAVPGATSYNIVISNGYKYEYFNTKSTATTW-----STKGKKIFPTNDB--IAK 534
Qy   474 GNCCITAQFDYVQPCFG---EVITRVKEKGAY--------AYI-------GSSP 509
Db   535 G--------BFEFHHDGKGTDFALDDPRAQYENAFQAGSTFGLRNLTRYLFRVQAVFPGESP 588
Qy   510 NSTWGEDYWSVGANAVFGVQP-------------------- 531
Db   589 NS----------DLVFAYMPIEKPQAPVAKAYSNLAHKETGVELNWEKSPMADGYKV 636
Qy   532 -TFEGTSMGSYDATFLEDSYNTVNSIMW---------AGNLAATHAGNIGNI-----TH 575
Db   637 LVFNGKAYEYDVG--AETKWTTQNKGIWPTKEETAEGKYALHHDGKGBAELAKDPSPVYTN 695
Qy   576 IGAHY------YW---EAYHVLGDG--SVMPYRAMPKTNTYTLPASLPQ--NQASY---- 618
Db   696 SGGNYKERTNYWFRVIAYQKAGNNTTSVQSEPATP-----SIPEALNKQLGMVDYWTSVP 750
Qy   619 ----SIQASAGSYVAISKDGVLYGTGVA--------NASGVA---TVSMTKQITEN 659
```

```
Qy    524  NAVFGVQPTFGTS--------------------------------MGSYDATFLEDS------------YNTVNSIM-  557
Db    134  NSL-PITDTINGTTELIIPTTSYNNQSHTLIYSTYTYTYLPNSTIDLSILPHSTISTLST  192

Qy    558  -----WAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNTYTLPASLP  612
Db    193  VSINDTSASLSKTTSPTAGTIT--------------------GYTSTFPASGT  233

Qy    613  QNQASYSIQASAGSYVAISKDCVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYL  672
Db    234  TSGTVEVVEPTAGTITETIVSGSVGYTSTFPANG--TTSGTVEVVEP------------TAGTIT  284

Qy    673  PVIKQIQVGEPSPYQPVSNLTATTQQCKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEP  732
Db    285  ETIVSGSVGYTSTFP-----ANGTTSG---TVEVVEPTAGTITETIVSGSVG--YTSTFP  334

Qy    733  ANDVRANEAKYV---------LAADNVWGDNTGYQFLLDADHNTFGS--VLPATGPLFTGTASS  785
Db    335  ANGTTSGTVEVVEPTAGTVTETIVSGSVGYTSTFPASGTTSGTVEVVEPTAGTITETIVS  394

Qy    786  NLYSANFEYLVPANADPVVTQNIIVTGQGEVVIPGGVGVYDYCITNPEPASG---KMWIAG  842
Db    395  G--SKAFTSTFPANG-------TT---SGTVEVV-------------EPTAGTITKTIVSG  430

Qy    843  DGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGL  902
Db    431  -------------------SKTFTSTFPANGTTSGT-VEVVEPTAGTITETIVSGSV---GY  469

Qy    903  TATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVIVEGSNEFAPVQNLTGSSVGQKVTL  962
Db    470  T-STFPANGTTSGTGTVEV--VEPTAG---TITETIVSGSKTFTSTFPASGTTSGTVEVV  521

Qy    963  KWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGHGWKPGNAPGIAGYNS  1022
Db    522  E---------------------PTAG-TITETIVSGSKAFTSTFPANGTTSGTVEVVEP-TAG--  561

Qy   1023  NGCVYSBSFGLGGIGVLT--PDNYLITPALDL--PNCGKLTFWVCAQDANYASEHYAVYA  1078
Db    562  ---TITETIVSGSVGYTSTFPASGTTSGTVEVVEPTAGTVTETIVSGSVGYTSTFPASGT  618

Qy   1079  SSTGNDASNFTNALLEETIT--AKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQ  1136
Db    619  TSGTVEVVEPTAGTVTETIVSGSVGYTSTFPASGTTSGT--VEVVEPTAGT-------  667

Qy   1137  STDMFYIDLDEVEIKANGKRADPTETFESSTHGEAPAEWTTIDADGDGQGWLCL--SSG  1193
Db    668  ------------ITETIVSGSKA-FTSTF------PANGTT------SGTVEVVEPTAG  701

Qy   1194  QLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMI  1253
Db    702  TITETIVSGSKTFTSTFPANGTTSG-----------TVEVVEPTAGTITETIVSGSVG--YTSTF  753

Qy   1254  SKTGTNAGDFTVVFEETPNGINK----GGARFGLSTEANGAKPQSVWIERTVDLPAGTKY  1309
Db    754  PASGTTSGTVEVV-EPTAGTVTETIVSGSVSFMSTIIAHDTSSGAVIV---VEPTAGTV-  808

Qy   1310  VAFRHYNCSDLNVILLDDIQF---------TMGGS------PTPTDYTYTVYRDGTKIKEGL  1356
Db    809  -------------TETIVSGSIPFTSTIPAQGTTSGTVEVVEPTAGTVTETIISGSV---GY  854

Qy   1357  TETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLT------AEQA  1409
Db    855  T-STFPAQGTTSGTVEV---VEPTAGTVTETIISGSVGYTSTFPAQSTTSGTVEMVAPTA  910

Qy   1410  PNSMDAILKWNAPASKRAEVLNEDFENGIPAS------------WKTIDADGDGN  1452
Db    911  GTVTETIVSGSIP-----------------FTSTIPAQGTTSGTVEVVEPTAGTVTETIISGSVG-  958

Qy   1453  NWTTTPPPGGSS------------SHINFEGPQNPDNYLVTPELSL  1499
Db    959  -YTSTFPAQGTTSGTVEIVAPTAGTVTETIVSGSIPFTSTIPAQGTTSG----TVEVVE  1012
```

RESULT 14
T39174
hypothetical Serine/threonine repeat containing protein [imported] - fission yeast (Schi
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence revision 09-Jun-2000 #text_change 09-Jun-2000
C;Accession: T50375; T39172; T39173, T39174; T39366
R;Connor, R.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21832
A;Accession: T50375
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1283 <CON>
A;Cross-references: EMBL:AL021815; PIDN:CAB61533.1; GSPDB:GN00067; SPDB:SPBC8E4.70c
A;Experimental source: strain 972h-; cosmid c8E4
A;Accession: T39172
A;Molecule type: DNA
A;Residues: 785-1283 <CO2>
A;Cross-references: EMBL:AL021815; PIDN:CAA17000.1
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
A;Accession: T39173
A;Molecule type: DNA
A;Residues: 'ME',179,'PLV',183,'W','KL',556-761,'HRGSS' <CO3>
A;Cross-references: EMBL:AL021815; PIDN:CAA17001.1
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
A;Accession: T39174
A;Molecule type: DNA
A;Residues: 785-1283 <CO2>
A;Cross-references: EMBL:AL021815; PIDN:CAA17000.1
A;Accession: T39173
A;Molecule type: DNA
A;Residues: 1-555,'S', <CO4>
A;Cross-references: EMBL:AL021815; PIDN:CAA17002.1
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Oliver, K.; Harris, D.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21848
A;Accession: T39366
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-896,'G',898-904,'I',906-1283 <WOO>
A;Cross-references: EMBL:AL035675; PIDN:CAB38695.1; GSPDB:GN00067; SPDB:SPBC1289.15
A;Experimental source: strain 972h-; cosmid c1289
C;Genetics:
A;Gene: SPDB:SPBC1289.15; SPDB:SPAC8E4.07c
A;Map position: 2
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

Query Match   2.9%; Score 265; DB 2; Length 1283;
Best Local Similarity   20.5%; Pred. No. 1.9e-06;
Matches 298; Conservative 174; Mismatches 535; Indels 448; Gaps 71;

```
Qy   1559  TRAQGTWYQKTVQLPAGTK--YVAFRHFGCTDFFWINLDDVVITSGNAPSYTYIYRNNT  1616
Db   1713  TRVND---------------------GSKNAYFEFDEFG-----NINVYTAGNGTAANYTYDSTQKVT  1754

Qy   1617  QIA-SGVTETTYRPDLATGFYTYGV--KVVYPN--GESAIETATLN-IT--SLADVT  1666
Db   1755  NAAISSASGTQILDEN------YTYDAASNRTSIDNKQDGKTTYEYDAVNQLIKETLPDGT  1809

Qy   1667  AQKPYTLTVVG-----------KTI--------TVTCQGEAMTYDMNGRRLAAGRNTV  1705
Db   1810  V-KAYTYDGFGNRTQVAISGSETKTIDASYNDGNQLVSWNGEALTYDANGNRTSDQKYT-  1867

Qy   1706  VYTAQGGHYAVMVVVDGKS  1724
Db   1868  -YTWDTGDRLSSITKKGES  1885
```

```
Qy    414  YTDVYNVLKAPVTGCYS--------HLNTGVSFANYTAHGSETAWADPLLTTSQLKAL  463
Db     70  YTTLITITNSQCTGNYSTLTISSQPLIHTNTSISKPSQTATFQNT-----NTTQV-SL  121

Qy    464  TNKDKYFLAIGNCCITAQFDVYQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGA  523
Db    122  TN--------------------------------GTTTNSY------SNT  133
```

```
Qy    1500 PGGGTLTFWVCAQDANYASEHYAV-YASSTGNDASNFANALLEVL-----------  1544           Qy    260 TLAAVSIGDVSSTNLKTRQLSGLAAGTSNTDAVNVAQLKVVDEIASRGWNLTASGAN--   316
           |||:|   |||     :                    ::                                  
Db    1013 PTAGTVTETIISGSVGTSTFPAQGTSTFGTVEVVEPTAGTVTETIVSGSIPFTSTIPAQG 1072          Db    583 EAYHVLGDGSVMPYRAMPKTNT---YTLPASLPQNQASYSIQAS----------AGSVAISKD  633

Qy    1545 ---TAKTVVTAPEAIRGTRAQGT-WTQKTVQLPA-GTKYVAFRHFGCTDFFWINLDDVV  1598          Qy    317 ------SGNVAPGSSVDLKNTDKNLTITKAIGSNDVQFNLNKDVKVTLAVGDALLNTD    369
           ||                                                                       
Db    1073 TTSGTVEIVVPTAGTVTAGTVTSTTGSGTSWFTTTV--PATGTR---------  1114          Db    634 GVLYGTGVA------------NASGV---------ATVSMTKQITENGNYDVVIT   667

Qy    1599 ITSGNAPSYTYTYIYRNNTQIASGVTETTYRDPDLATGFYTYGVKV-------  1649          Qy    370 GIALGTDVSLSTGLAITDGPAVTASGIDAGSKVISHVAAGAVSETSTDAVNGSQLNAV-   428
           :|    ||                                                                  
Db    1115 VVSPTAPACSSTFFA------------ECPNLNFPHSKNVNFPPADIRVISVTPKGN    1160          Db    668 RSNYLPVIKQIQVGEPSYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVKRIGDGLF   727

Qy    1650 SAIETATLNITSLADVTAQKPYTLJTVVGKITITVCQGEAMIYDMNGRRLAAGRNTVVTA 1709          Qy    429 -------QVQASQPVTFGNEGAVKRSLGQSVVISGDSSSTAGTYSGG-NLKSVVD--    475
           |:     :|   |:|: |||                                                     
Db    1161 NMYD-ATVQFTTSSTMSKKSLSELKILGLSQTY------LLIYSYNSK-------  1205          Db    728 VTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTG-TASSN  786

Qy    1710 QGGHYAVMVVVDGKS                                               1724          Qy    476 ------EAAGRIHLQLADSPKFGNVVINNGGKISGVTAGTE                     510
           |||                                                                      
Db    1206 NPGSWTSTVTVQGTS                                               1220          Db    787 LYSA-NFEYLVPANADPVVTTQNIIVT---GQGGVVIPGGVGYDYCITNPEPASGKMWIAG  842

                                                                                      Qy    511 ETDAVNFSQL---KSISTAVDGWTLTASGSNGSKVASGGTVDLKNTD---------GNLTISK  562
RESULT 15                                                                             
D82671                                                                         Db    843 DGGNQPARYD---DFTFEAGKKYTFTMRRAGMGDGTDMEVE---------DDSPASYTYTYVR  893
surface protein XF1529 [imported] - Xylella fastidiosa                        
C;Species: Xylella fastidiosa                                                 Db    563 SGDSNDVVFNLSKDFKVDGMTSGTTVVNNDGVKVGSDVALGTTGLTITDGPA-VTASGID   621
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000   
C;Accession: D82671                                                            Qy    894 DGTKIKEGLTATIFEEDGVAA--GNHEYCVEVKT----------AGVSPKVCKDVTVEGS   942
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
submitted to GenBank, June 2000                                               Db    622 AGSKVISHVAAGVVSETSTDAVNGSQLNAVQVQASQPVTFGNEGAVKRSLGQSVVISGE   681
Nature 406, 151-157, 2000                                                     
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.        Qy    943 NEFAPVQN------------LTGSSVGQKVTLKWDAP                          967
A;Reference number: A82515; MUID:20365717; PMID:10910347                      
A;Note: for a complete list of authors see reference number A59328 below      Db    682 SSTAGMYSGGNLKSVVDEAAGRIHLQLADSPKFGNVVINNGGKISGVTAGTEET---DAV   738
A;Accession: D82671                                                           
A;Status: preliminary                                                         Qy    968 NGTPNPNPNPNPNPGTTLSESFENGIP-ASWKTID---ADGDGHGWKPGNA---------- 1014
A;Molecule type: DNA                                                          
A;Residues: 1-2059 <SIM>                                                      Db    739 NFSQLKSISTAVDGWTLTASGSNGSKVASGGTVDLKNTDGNLTISKSGDSNDVVFNLSK   798
A;Cross-references: GB:AE003849; NID:g9106554; PIDN:AAF84338.1; GSPDB:GN001    
A;Experimental source: strain 9a5c                                            Qy   1015 -------PGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLITPWVCA--  1064
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  Db   799 DFKVDGMTSGTTVVNNDGVKVGSDVALGTTGLTITDGPAVT-ASGIDAGSKVISHVAAGV  857
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.      
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  Qy   1065 ----QDANYASEHYAVYASS----------TGNDASNFTNALLEETITAKGVRSPKAIRGRIQ 1113
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Laigr     
chado, M.A.; Madeira, A.M.B.N.; Marino, C.L.; Marques, M.V.; Martins, E        Db   858 VSETSTDAVNGSQLNAVQVQASQPVTFGNEGA--VKRSLGQSVVISGESS----TAGTYS   912
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A   Qy   1114 GTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADF---TETFESSTHGE  1170
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silvei   Db   913 GGNLKSVVDEAAGR-----IHLQLADS--PKFGNVVINNGGKISGVTAGTEETDAVNFSQ   965
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328                                                     Qy   1171 APAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGS----------NVVSFSWNGMALNPD-- 1220
A;Contents: annotation                                                        
C;Genetics:                                                                   Db   966 LKSISTAVD-------QGWLTASGANGSKVASGGTVDLKNTDGNLTISKSGDSNDVVFN-- 1018
A;Gene: XF1529                                                                 
                                                                              Qy   1221 NYLISKD--------VTGATKVKYYYAVNDGFP-GDHYAVMISKTGTNAGDFTVVFEETPNG 1273
Query Match    2.9%;   Score 262;  DB 2;  Length 2059;                         
Best Local Similarity  19.1%;                                                 Db   1019 --LSKDFKVDGMTSGTTVVN----NDGVKVGSD--VALGTGLTIANGPAV---TASG    1065
    Pred. No. 6.1e-06;                                                        
Matches 319;  Conservative 197;  Mismatches 570;  Indels 580;  Gaps 74;        Qy   1274 INKG--------GARFGLSTEA-NGAKPQSVWIERTVDL--------               1303
                                                                              
Qy     403 GMCQYYYNCEHGYTDVYNYLKAPYTGCYSHLNTGVSFANYTAHGS-ETAWADPLLTTSQLK 461           Db   1066 IDAGSKVISHVAAGAVSETSTDAVNGSQLNAVQVQASQPVTFTGNEGAVKRSLGQSVVIS  1125
           ::      :                 :||                 :||
Db     103 GSLYVNSCQVGVNDVLNKTYSIRMGSVITMNT-VAGTNAIAIGSAQSSAADALKASLATK  161           Qy   1304 PAGTKYVAFRHYNCSDLNYILLDD--------IQFTMGGSPTPTDYTYTYRDGTKIKEG   1355

Qy     462 ALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSV  521           Db   1126 YSGGNLKSV-VDEAAGRIHLQLADSP-----KFGNVVINNGGKI-SG              1173
           ::                               :
Db     162 A---SGARAIAIG----AKASADGVD-------TVALGSGATAGTGASSIAI------GL  201           Qy   1356 LTETTFEEDG----------VATGNHEYCVEVKYTAGVSPKK                    1387

Qy     522 GANAVFGV------------QPTFEGTSMGSYDATFLEDSYNTVNSIMWAG           560           Db   1174 VTAGTEETDAVNFSQLKSISTAVDQGWTLTASGANGSKVASGG---TVDLKNTDG-----  1225
           |||
Db     202 NASAVNGAVAVGGGALVTVPDGAVALGLNSVASTGKGLSGYDPK--TKTTSTDASAAWKS  259           Qy   1388 CVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFE-NGIPASWKTID  1446

Qy     561 NLAATHAGNIG------------------NITHIGAHYW                       582           Db   1226 -------NLTISKSGDSNDVVFN----------LSKDFKVDGMTSGTTVVN           1259
           |||                            ||:
```

```
Qy    1447 ADG-----------------DGNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPD 1489
                               :  :||       : :|
Db    1260 NDGVKVGSDVALGTTGLTIANGPAVTASGIDAGSKVISHVAAGAVSETSTDAVNGSQ--- 1316

Qy    1490 NYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTV 1549
                  : :   | |  : |       ||| :    ::
Db    1317 LNAVQVQASQP----VTF-----------TGNEGA-------VKRSLGQSVV 1346

Qy    1550 VTAPEAIRGTRAQGTWYQKTVQLPAGTKVVAFRHFGCTDFFWINLDDVVITSGNAPSYTY 1609
           ::  :||:  :  :|||          : : ::|:
Db    1347 ISGESSTAGTYSGGN-LKSVVDEAAGT---------------IHLQLADSPKFGN 1385

Qy    1610 TIYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVT--A 1667
                :  |||||           |       :  :  :   ::
Db    1386 VVINNGGKI-SGVTAGTE---------------------ETDAVNFSQLKSISTAV 1419

Qy    1668 QKPYTLTVVGK----------------TITVTCQGEA-------MIYDMNGRRLAAGR- 1702
              : :||||                :|::  |::       :   :  |: |
Db    1420 DQGWTLTASGANGSKVASGGTVDLKNTDGNLTISKSGDSNDVVFNLSEDLKEKSITVGNT 1479

Qy    1703 ----------------NTVVYTAQGGHYAVMVVVDGKSYVEKLAV 1731
                           :|: :|  : :|:|| |  |: |:|
Db    1480 QLDKDGVKVSSNVLLDSNELVITSHSSTSSVKTLANGESVVNRTVV 1525
```

Search completed: May 18, 2004, 11:36:13
Job time : 48.8582 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:22:23 ; Search time 21.7568 Seconds
                              (without alignments)
                              4145.156 Million cell updates/sec

Title: US-08-353-485-10
Perfect score: 9179
Sequence: 1 MRKLLLLIAASLLGVGLYAQ.........HYAVMVVVDGKSYVEKLAVK 1732

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	4217	45.9	2628	1	HGA2_PORGI	Q51845 porphyromon
2	4200.5	45.8	2164	1	HGA_PORGI	P59915 porphyromon
3	1946.5	21.2	989	1	PRTH_PORGI	P46071 porphyromon
4	1222.5	13.3	991	1	CPG1_PORGI	P28784 porphyromon
5	541	5.9	736	1	CPG2_PORGI	P95493 porphyromon
6	250.5	2.7	1200	1	HYAL_STRPU	O76536 strongyloce
7	242.5	2.6	2334	1	WAPA_BACSU	Q07833 bacillus su
8	224	2.4	2894	1	YD96_METJA	Q58791 methanococc
9	223	2.4	1541	1	IGA1_HAEIN	P42782 haemophilus
10	220.5	2.4	1902	1	P2P_LACLC	P15293 lactococcus
11	219.5	2.4	1953	1	BIGA_SALTY	P25927 salmonella
12	218.5	2.4	3178	1	YS89_CAEEL	Q09624 caenorhabdi
13	214	2.3	1861	1	APU_THETU	P38536 t amylopull
14	211	2.3	2358	1	YEEJ_ECOLI	P76347 escherichia
15	209.5	2.3	1655	1	OMPB_RICCN	Q9kka3 r outer mem
16	209	2.3	1090	1	GUXB_CELFI	P50899 cellulomona
17	209	2.3	2660	1	YEEJ_ECO57	Q8x8v7 escherichia
18	208.5	2.3	1902	1	P1P_LACLC	P16271 lactococcus
19	202.5	2.2	1481	1	APU_THEET	P38939 t amylopull
20	202.5	2.2	1656	1	OMPB_RICJA	O06653 r outer mem
21	202.5	2.2	1848	1	CBPA_CLOCL	P38058 clostridium
22	200.5	2.2	1654	1	OMPB_RICRI	Q53047 r outer mem
23	199	2.2	1645	1	OMPB_RICTY	P96989 r outer mem
24	196.5	2.1	1902	1	P3P_LACLC	P15292 lactococcus
25	194	2.1	1279	1	APU_THESA	P36905 t amylopull
26	194	2.1	2249	1	OMPA_RICRI	P15921 rickettsia
27	192	2.1	1475	1	APU_THETY	P16950 t amylopull
28	187	2.0	1243	1	VG37_BPK3	Q38394 bacteriopha
29	187	2.0	1256	1	ATL_STAAU	P52081 staphylococc
30	185	2.0	1902	1	P2P_LACPA	Q02470 lactobacill
31	184.5	2.0	1770	1	PMPC_CHLTR	O84419 chlamydia t
32	184	2.0	1450	1	MPSF_CHICK	Q02113 gallus gall
33	183.5	2.0	1260	1	ALS1_CANAL	P46590 candida alb
34	183.5	2.0	1577	1	HLYA_PROMI	P16466 proteus mir
35	183	2.0	2021	1	OMPA_RICCN	Q52657 rickettsia
36	181	2.0	4590	1	FATH_HUMAN	Q14517 homo sapien
37	179.5	1.9	1325	1	120K_RICRI	P14914 rickettsia
38	177.5	1.9	1341	1	YDEK_ECOLI	P32051 escherichia
39	175.5	1.9	1829	1	VG37_BPT2	P07067 bacteriopha
40	175.5	1.9	4303	1	FRPC_NEIMB	Q9jyv5 neisseria m
41	175.5	1.9	6632	1	PKD1_HUMAN	P98161 homo sapien
42	175	1.9	872	1	UN89_CAEEL	O01761 caenorhabdi
43	174.5	1.9	1419	1	GUXA_CELFI	P50401 cellulomona
44	174.5	1.9	1045	1	ALA1_CANAL	O13368 candida alb
45	174	1.9		1	GUNB_CELFI	P26225 cellulomona

ALIGNMENTS

RESULT 1
HGA2_PORGI
ID   HGA2_PORGI   STANDARD;   PRT;   2628 AA.
AC   Q51845;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97047672; PubMed=8926061;
RA   Han N., Whitlock J., Progulske-Fox A.;
RT   "The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381
RT   contains four large, contiguous, direct repeats.";
RL   Infect. Immun. 64:4000-4007(1996).
CC   -!- FUNCTION: Agglutinates erythrocytes
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U41807; AAB17128.1; -.
DR   PIR; T28651; T28651.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT   SIGNAL   1   24   POTENTIAL.
FT   CHAIN   25   2628   HEMAGGLUTININ A.
FT   DOMAIN   540   995   PEPTIDASE C25-LIKE 1.
FT   DOMAIN   996   1451   PEPTIDASE C25-LIKE 2.
FT   DOMAIN   1452   1907   PEPTIDASE C25-LIKE 3.
FT   DOMAIN   2074   2628   PEPTIDASE C25-LIKE 4.
FT   DOMAIN   PEPTIDASE C25-LIKE 5.
SQ   SEQUENCE   2628 AA; 283324 MW; 61C4DB32540C99DA CRC64;

Query Match   45.9%; Score 4217; DB 1; Length 2628;
Best Local Similarity   51.0%; Pred. No. 4.1e-223;
Matches 958; Conservative 114; Mismatches 363; Indels 442; Gaps 53;

Qy   179 YDVVANQLKVRNNIBIEVSFQ-----GADEVATQRLYDASFSPYFETAYKQLFNRDVYT 232
                                        ||: : :: : : :||::
Db   871 YTVYRDGTKIKEGL-TETTFEEDGVATGNHEYCVEVKYTAGVSP-----KECVNVTV-- 921

Qy   233 DHGDLYNTPVRMLIVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIH 292
                     ||:|| :|:::| : | :| ||||
Db   922 --DPVQFNPVQNLTGSAVGQKVTLK----WDAPNG---TPNPNPNPGTTTLS------- 966

```
Qy   293  KYNDGLAAS--------------------------AAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAV   336
          : :: :.|                            |||||||||||||||||||||  :|:|||
Db   967  ESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSAS--------------------YINF   1017

Qy   337  DGDYFPEMYTFRMSASSP--------------BELTNIIDKVLMYEKATMPDKS-----YLEK   380
          ||||:|                          |:: |:  :|                |||||
Db  1018  EGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEE   1077

Qy   381  VL--------LIAG--ADYSWNSQVGQPTIKYGMQYYYNQEHGYYDVY----   418
          ||         |||.  |.:||.  |: :: :||  ::: :: :||
Db  1078  VLTAKTVVTAPEAIRGTRVQGTWYQKTVQ--LPAGTKYVAFRHFGCTDFFWINLDDVEIK   1135

Qy   419  -NYLKAPYTGCYSHLNTGVSFANYT--AHGSETAWADPLLTTSQLKALTNKDKYFLAIG   474
          :||  ||    :  : ||:| :|     |::  | :. |||: :  |||   |:: ||
Db  1136  ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL--CLSSGQLGWLT-----AHG   1186

Qy   475  NCCITAQFDY-----VQP-----CFGEVITRVKEKGAYAYIGSSPNSYWGEDYWSVGANAV   526
          ||:   |:|      | |      |   : :|.  |: ||:: .| :|   ||:: :|| |
Db  1187  GTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYYAVNDGFPGDH--------------   1232

Qy   527  FGVQPTTEGTSWGSYDATFLEDSYNTVN----------SIMWAGNLAATHAGN   569
          ||   |.|||.:|: :|:|   |::| |          ||  ::| |: ||
Db  1233  YAVMISKTGTNAGDFTVVF-ETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGT   1291

Qy   570  IGNITHIGAHY--WEAYHVLGDGSVMPYRAMPKTNTYTLPASLPQNQASYSIQASAGSY   627
          :||:  |   |   : |:.: .   |  : :: |:|| ||:|  |||| |: ||||:|
Db  1292  ---KVVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT------------------   1326

Qy   628  VAISKDGVLYGTGVANAS----GVATVSMTKQITENGNYDVVITRSNYLPVIKQIVQVG-   681
          |||:|||.||||  ||:        |:|  |||    |  :  |:||||  |:|:|||.
Db  1327  YTVYRDGTKIKEGLTETTFEEDGVAT---------GNHEYCVEVKYTAGVSPKECVNVT   1376

Qy   682  -EPSPYOPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPANDVRANE   740
          :  |:| |: .|  |   :||||: |.  .|  :  |.  :  ||:|| ::.|||: ||:
Db  1377  VDPVQFNPVQNLTGSAVGQKVTLKWDAPNG------------------TPNPNPNPNP   1416

Qy   741  AKVVLAADNWGDNTGYQFL-LDADHNTFGSVIPATGPLFTG-----TASSNLYSANFE-   793
          || :   :|||.||||||   |: :|  ||:|  | |||        |||| ||||
Db  1417  GTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASY-INFEG   1475

Qy   794  -------YLVP----------------ANA-DPVV   804
                 |:||                ||| :|||
Db  1476  PQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVL   1535

Qy   805  TTQNII--------VTG--QGEVIPGG----------   822
          |: :||        |||  :|.|:|.:
Db  1536  TAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGK   1595

Qy   823  -----------VYDYCITNPEPA--------------SGKM-WIAGDGGN-   846
                     || : :|.|.||            |||:  ||  |.|
Db  1596  RADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWNGM   1655

Qy   847  --QPARY-------------------DDFT--FE---   857
            :| |||                  :||   ||
Db  1656  ALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGIN   1715

Qy   858  --------------------AGKKYTFTMRRAGMGDGTDMEVED--   881
                              |  |: .| .: |.|||.||  .
Db  1716  KGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYV-APRHYNCSDLNYILLDDIQFTMGG   1774

Qy   882  -DSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVE   940
           :|.||:|||||||||||||||:|||||||||:|||||||||||||||||  | |||| |
Db  1775  SPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVD   1834

Qy   941  GSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTILSESFENGIPASWKTI   1000
          ||||.||||||||| ||||||||||||| :|::  || || ||||||||||||||||||
Db  1835  -PVQFNPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGT-TTLSESFENGIPASWKTI   1892

Qy  1001  DADGDGHGWKPGNAPG--IAGYNSNGCVYESFGLGGIGVLTPDNYLITPALDLPNGGK   1057
          ||||||.||.||||    ||.|||.||||||| |  |.|||||||:|||| |||||||
Db  1893  DADGDGNNWTTTPPGGTSFAGHNSAICVSSASY-INFEGPQNPDNYLVTPELSLPGGGT   1951

Qy  1058  LTFWVCAQDANYASEHYAVYASSTGNDASNFNALLEETITAKG-VRSPKAIRG-RIQGT   1115
          ||||||||||||||||||||||||||||||||  |||:|:||   ||||.|:|  |:||
Db  1952  LTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGT   2011

Qy  1116  WRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEW   1175
          |.|||| |||||||||||||| ||:||::|||:|||||||||||||||||||||||||||
Db  2012  WYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIKANGKRADFTETFESSTHGEAPAEW   2071

Qy  1176  TTIDADGDGQWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVK   1235
          |||||||||.||||||||| |||||| ||| ||| |||||||||||||||||||||||||
Db  2072  TTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVK   2131

Qy  1236  YYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSV   1295
          |||||||||||||||||||||| |||||||||||||||||||||||||||||.||||||
Db  2132  YYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEADGAKPQSV   2191

Qy  1296  WIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEG   1355
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2192  WIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEG   2251

Qy  1356  LTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDA   1415
          |||||||||||||||||||||||||||||||  |||||| |: |||||||||:|||||||
Db  2252  LTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTVINPTQFNPVQNLTAEQAPNSMDA   2311

Qy  1416  ILKWNAPASKRAEVLNEDFENGIPASWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVS   1475
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2312  ILKWNAPASKRAEVLNEDFENGIPSSWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVS   2371

Qy  1476  SASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF   1535
          || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2372  SASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF   2431

Qy  1536  ANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFPGCTDFFWINLD   1595
          ||||||||||||||||||||||||| ||||||||||||||||||||||| ||||||||||
Db  2432  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD   2491

Qy  1596  DVVITSGNAPSYTYIYRNNTQIASGVTETTYRDPDLIATGFYTYGVKVVYPNGESAIETA   1655
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2492  DVVITSGNAPSYTYTYRNNTQIASGVTETTYRDPDLIATGFYTYGVKVVYPNGESAIETA   2551

Qy  1656  TLNITSLADVTAQKPYTLITVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYA   1715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2552  TLNITSLADVTAQKPYTLITVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYA   2611

Qy  1716  VMVVVDGKSYVEKLAVK   1732
          |||||||||||||||||
Db  2612  VMVVVDGKSYVEKLAVK   2628
```

```
RESULT 2
HGA1_PORGI
ID   HGA1_PORGI              STANDARD;      PRT;   2164 AA.
AC   P59915;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HGA OR PG1837.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=22829867; PubMed=12949112;
RA   Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA   Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
RA   Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
RA   Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA   Dewhirst F.E., Fraser C.M.;
RT   "Complete genome sequence of the oral pathogenic bacterium
RT   Porphyromonas gingivalis strain W83.";
RL   J. Bacteriol. 185:5591-5601(2003).
```

```
CC   -!- FUNCTION: Agglutinates erythrocytes (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AE017178; AAQ66831.1; ALT_INIT.
DR   TIGR; PG1837; -.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
KW   Complete proteome.
FT   SIGNAL        1     25                    POTENTIAL.
FT   CHAIN        26   2164                    HEMAGGLUTININ A.
FT   DOMAIN       26    539                    PEPTIDASE C25-LIKE 1.
FT   DOMAIN      540    991                    PEPTIDASE C25-LIKE 2.
FT   DOMAIN      992   1443                    PEPTIDASE C25-LIKE 3.
SQ   SEQUENCE   2164 AA;  233387 MW;  6DFAB22832586C63 CRC64;

Query Match             45.8%;  Score 4200.5;  DB 1;  Length 2164;
Best Local Similarity   54.4%;  Pred. No. 2.5e-222;
Matches   924;  Conservative 115;  Mismatches 320;  Indels 338;  Gaps   45;

Qy          295 YNDGLAASAAPVFLALVG------DTDVISGEKGKKITKKVTDLYYSAVDGDYFPEMYIFRMS 350
                                        :|:                  :  :
Db          547 HNSAICVSSAS-YINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAV-YA 604

Qy          351 ASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAG--ADYSWNSQVGQPTIKYGMQYYY 408
                :|:.||                  | |                |  || | || ||||
Db          605 SSTGNDASNFANALL---EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQ--LPAGTKYVA 659

Qy          409 NQEHGYTDVY---------NYLKAPYTGCYSHLNTGVSFANYT---AHGSETAWADPL 454
                                    :  |  | : | || |      |  ||
Db          660 FRHFGCTDFFWINLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL---C 717

Qy          455 LTTSQLKALTNKDKYFLAIGNCCITAQFDY-----VQP---CFGEVITRVKEKGAYAYIG 506
                ||::                          :      |          ::: |
Db          718 LSSGQLGWLT-----AHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVND 770

Qy          507 SSPNSYWGEDYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVN---------- 554
                                        :| |  |    ||| :||::: |||
Db          771 GFPGDH---------------YAVMISKTGTNAGDFTVVF-EETPNGINKGGARFGLSTEA 815

Qy          555 -----SIMWAGNLAATHAGNIGNTHIGAHYY--WEAYHVLGDGSVMPYRAMPKINTYTL 607
                     :||:  |     | :|     |  ||   |  : | ::
Db          816 NGAKPQSVWIERTVDLPAGT--KVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT- 870

Qy          608 PASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANAS---GVATVSMIKQITENGNYD 663
                :||:||  ||    ||| || : ||  |  ||  ||          || | :||||||
Db          871 ---------------YTVYRDGTKIKEGLTETTFEEDGVAT---GNHE------------- 900

Qy          664 VVITRSNYLPVIKQIQVG--EPSPYQPVSNLTATTQGQKVTLKWEAPS------------ 709
                :|:  :  | | | || |     | |  | ||    ||:||:||
Db          901 YCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPNGTPNPNPGTT 960

Qy          710 -AKKAEGS-----REVKRIGDGLFVIEP----ANDVRANEAKVVLAAD--NVWGDNTGY 757
                          :|: ||                 |  | : | ||||  ||  || :
Db          961 LSESFENGIPASWKTIDADGDGNNWTTPPPGGSFAGHNSAICVSSASYINFEGPQNPD 1020

Qy          758 QFLL---------------DADHNTFGSVIPATGPLFTGTASSNLYSANFEYLIVPA 798
                :|                  |  | || |:||| | || :|| ||  |: :|::||
Db         1021 NYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASS---TGNDASNFANALLEEVLTA 1077

Qy          799 NADPVVTTQNII----VTG--QGEVIPGG----------------------------- 822
                :|  |  | :      :|   :|| |  |
Db         1078 KT--VVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGK 1135

Qy          823 --------VYDYCITNPEPA------------SGKM-WIAGDGGN------------- 846
                         :| |   :|           || : | ||   |
Db         1136 RADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASPSWNGM 1195
```

```
Qy          847 --QPARY--------------------------------------------DDFT--FE------ 857
                                                                   |||  ||
Db         1196 ALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEETPNGIN 1255

Qy          858 ------------------------------AGKKYTFTMRRAGMGDGTMEVED-------- 881
                                              ::  :
Db         1256 KGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYV-AFRHYNCSDLNYILLDDIQFTMGG 1314

Qy          882 -DSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVE 940
                 : |||||||||||||||||||: || :||||| |||||||||||||| | ||  |: |
Db         1315 SPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVD 1374

Qy          941 GSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTI 1000
                  :  |||||||||:|||||||||||||||:||||| |      ||||||||||||||||
Db         1375 -PVQFNPVQNLTGSAVGQKVTLKWDAPNGTPNPNPGT-----TTLSESFENGIPASWKTI 1428

Qy         1001 DADGDGHGWKPGNAPG---IAGYNSNGCVYSESFGLGGIGVLITPDNYLITPALDLPNGGK 1057
                |||||| || ||| |      |  :|| |::  |: |     ||||||:  | |||||||
Db         1429 DADGDGNNWTTPPPGGTSFAGHNSAICVSSSASY-INFEGPQNPDNYILVTPELSLPNGGT 1487

Qy         1058 LTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETIATAKG-VRSPKAIRG-RIGT 1115
                ||||||||||||||||||||||||||||||:||| ||||::: ||  :| ||||| |:||
Db         1488 LTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGT 1547

Qy         1116 WRQKTVDLPAGTKYVAFREFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEW 1175
                | |||| |||||||||||| | :     :  |:|||||||||||||||||||||||||||
Db         1548 WYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGKRADFTETFESSTHGEAPAEW 1607

Qy         1176 TTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVK 1235
                ||||||||||||||||||||||||||||:||:|:|| |||||||||||||||||||||||
Db         1608 TTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVK 1667

Qy         1236 YYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSV 1295
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1668 YYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSV 1727

Qy         1296 WIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEG 1355
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1728 WIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEG 1787

Qy         1356 LTETTFEEDGVATGNHEYCVEVKYTAGVSPKCVDVTVNSTQFNPVQNLTAEQAPNSMDA 1415
                ||||||||||||||||||||||||||||| ||||||:|: |||||||||:||||||:|||
Db         1788 LTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVNPTQFNPVQNLTAEQAPNSMDA 1847

Qy         1416 ILKWNAPASKRAEVLNEDFENGIPASWKTIDADGDGNNWTTPPPGGSSFAGHNSAICVS 1475
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1848 ILKWNAPASKRAEVLNEDFENGIPASWKTIDADGDGNNWTTPPPGGSSFAGHNSAICVS 1907

Qy         1476 SASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 1535
                || :||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1908 SASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 1967

Qy         1536 ANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 1595
                ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db         1968 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 2027

Qy         1596 DVVITSGNAPSYTYTYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPNGESAIETA 1655
                || :::||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2028 DVVITSGNAFSYTYTYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPNGESAIETA 2087

Qy         1656 TLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYA 1715
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         2088 TLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYA 2147

Qy         1716 VMVVVDGKSYVEKLAVK 1732
                ||||||||||||||||
Db         2148 VMVVVDGKSYVEKLAVK 2164
```

```
RESULT 3
PRTH_PORGI
ID   PRTH_PORGI   STANDARD;   PRT;   989 AA.
```

```
AC   P46071;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protease prtH (EC 3.4.22.-).
GN   PRTH.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=95012612; PubMed=7927685;
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RT   "Cloning and characterization of a new protease gene (prtH) from
RT   Porphyromonas gingivalis.";
RL   Infect. Immun. 62:4279-4286(1994).
RN   [2]
RP   ERRATUM.
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RL   Infect. Immun. 62:5707-5707(1994).
CC   -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
CC       P.GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
CC       IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
CC       AND IS A VIRULENCE FACTOR.
CC   -!- SUBCELLULAR LOCATION: In membrane vesicles.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L27483; AAA51298.1; -.
DR   HSSP; P23882; 1FMT.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR002376; formyl_transf.
DR   Pfam; PF00551; formyl_transf; 1.
KW   Hydrolase; Thiol protease; Repeat; Virulence.
FT   REPEAT      270    323
FT   REPEAT      528    581
SQ   SEQUENCE   989 AA;   110238 MW;   FA85FEBA3AC8944C CRC64;

Query Match             21.2%;       Score 1946.5;  DB 1;  Length 989;
Best Local Similarity   61.3%;       Pred. No. 3.9e-99;
Matches  411;  Conservative  45;  Mismatches  125;  Indels  89;  Gaps  13;

Qy 1100 KGVRSPKAIRGRIQGTWRQKTVDLPAGTKVYAFRHFQSTDMFYID-LDEVEIKANGKRAD 1158
        |||||| :||||||||||||||||||||||       :|:  |:|  |  :| :|||||||
Db   32 KGVRSPEAIRGRIQGTWRQKTVDLPAGTEICCFPSLPKAPICSTSTSTMRLRSKTNAKRAD 91

Qy 1159 FTETFESSTHGEAPAEWTTIDADGDQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALN 1218
        |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   92 FTETFESSTHGEAPAEWTTIDADGDQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALN 151

Qy 1219 PDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGG 1278
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  152 PDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGG 211

Qy 1279 ARFGLSTEANGAKPQSVWIERTVDLPAGTKVYAFRHYNCSDLNYILLDDIQFTMGGSPTP 1338
        |||||||||:|||||||||||||||||||||||||||||||||||||||||:|||||||||
Db  212 ARFGLSTEADGAKPQSVWIERTVDLPAGTKVYAFRHYNCSDLNYILLDDIQFTMGGSPTP 271

Qy 1339 TDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQF 1398
        ||||||||||||||||||||||||||||||||||||||||||||||||||:||:|||||||
Db  272 TDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTVNSTQF 331

Qy 1399 NPVQNLTAEQAPNSMDAILKWNAPASKRAE----------------------------V 1429
        |||:||| |: |: ||:|||||||| |: |:
```

```
Db  332 NPVKNLKAQ--PDGGDVLKWEAPSAKKTEGSREVKRIGDGLFVTIEPANDVRANEAKVV 389
Qy 1430 LNEDFENGIPASWKTIDADGDGNNWTTPPGGSSFAGHNSAICVSSASHINFEG--PQN 1487
Db  390 LAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSDLYSA---NFEYLIPAN 444
Qy 1488 PD-----NYLVT--PELSLPGGGTLTFWVCAQDANYASEHYAYYASSTGNDASNFANAL 1539
        ||     :||::  |: ||||
Db  445 ADPVVTTQNIIVTGQGEVVIPGG---VYDYCITNPEPASGKMWI-AGDGGNQPARYDDFT 500
Qy 1540 LEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKVV-AFRHFGCTDFFWINLDDVV 1598
        |
Db  501 FE----------------------------AGKKYTFTMRRAGMGDG-----TDME 523
        |:                             :
Qy 1599 ITSGNAPSYTYTYRNNTQIASGVTETYRDPDLIATGFYTYGVKVVYPNGESAIETATLN 1658
        :||| :||:|||||:|::|:|||: |:||:   |||||::||||::|
Db  524 VEDDSPASYTYTYRDGTKIKEGLTETYRDAGLSAQSHEYCVEVKYTAGVSPKVCVDYI 583
Qy 1659 ITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVVTAQGHYAVMV 1718
        |||| ||||||||||||||||||| ||||||||||||||||||||||||||||||:|||||
Db  584 PDGVEDVTVQKPHTLTVVGKTITVSWQGEAMIYDMNGRRLAAGRNTVVVTAQGGYAVMV 643
Qy 1719 VVDGKSYVEK 1728
        |||||||||
Db  644 VVDGKSYVEE 653
```

```
RESULT 4
CPG1_PORGI
ID   CPG1_PORGI    STANDARD;    PRT;    991 AA.
AC   P28784; Q45168;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
DE   (RGP-1).
GN   RGPA OR RGP1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 228-290 AND 517-541.
RC   STRAIN=381;
RX   MEDLINE=95168884; PubMed=7864651;
RA   Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
RA   Ikehara Y.;
RT   "Structural characterization of argingipain, a novel
RT   arginine-specific cysteine proteinase as a major periodontal
RT   pathogenic factor from Porphyromonas gingivalis.";
RL   Arch. Biochem. Biophys. 316:917-925(1995).
RN   [2]
RP   SEQUENCE OF 228-270.
RC   STRAIN=HG66;
RX   MEDLINE=92406812; PubMed=1527017;
RA   Chen Z., Potempa J., Polanowski A., Wikstrom M., Travis J.;
RT   "Purification and characterization of a 50-kDa cysteine proteinase
RT   (gingipain) from Porphyromonas gingivalis.";
RL   J. Biol. Chem. 267:18896-18901(1992).
CC   -!- FUNCTION: Thiol protease which is believed to participate in
CC       intracellular degradation and turnover of proteins. Its
CC       proteolytic activity is a major factor in both periodontal tissue
CC       destruction and in bacterial host defense mechanisms. Activates
CC       complement C3 and C5.
CC   -!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
CC       immunoglobulins, with a preference for Arg in P1, and hydrophobic
CC       residues in P2 and P3.
CC   -!- ENZYME REGULATION: Requires cysteine for activation and Ca(2+)
CC       and/or Mg(2+) for stabilization. It is stimulated by glycine-
CC       containing dipeptides. It is resistant to inhibition by proteinase
CC       inhibitors in human plasma.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   -----------------------------------------------------------------------
```

```
DR   EMBL; D26470; BAA05484.1; -.
DR   PIR; I40229; I40229.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase C25.
DR   InterPro; IPR005536; Peptidase C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
KW   Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
FT   SIGNAL        1     24
FT   PROPEP       25    227
FT   CHAIN       228    991       GINGIPAIN R1.
FT   ACT_SITE    438    438       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    471    471       NUCLEOPHILE (BY SIMILARITY).
FT   CONFLICT    264    265       RT -> TK (IN REF. 2).
SQ   SEQUENCE   991 AA;  108782 MW;  03EE3F43CEBE2544 CRC64;
```

```
Query Match             13.3%;  Score 1222.5;  DB 1;  Length 991;
Best Local Similarity   31.8%;  Pred. No. 2.2e-59;
Matches  322;  Conservative 174;  Mismatches 381;  Indels 135;  Gaps 32;

Qy       6 LLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSPKQFDASFSFNEVELTKVETKGGTFA 65
             :: |||||  :  :    :: ||:|:|   :      ||:| : | :|:|
Db      10 IALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGMAQ 67

Qy      66 SVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKS--FTEQVYSLNQYGSEKLMPH 123
             :|  :   || |  || ||  |    |    ||| ||       |||:|:|  || |||
Db      68 VPTYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL------IAPS 120

Qy     124 QPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYDVVA 183
             : :|:   || ||:  :  :|   || :|: ||||:| | || ||  | ||:  || |
Db     121 KGMIMRNEDDPKKIPYVY-GKSYSQNKFFPGEIATLDDPFLLRDVRGQVVNFAPLQYNPVT 179

Qy     184 NQLKVRNNIEI--EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
              :|  : :    :   : || :   :|:        |    :|:          |  |:
Db     180 KTLRIYTEITVAVSETSEQGKNILNKKGTFAG------FEDTYKRMF------MNYEPGRYT 229

Qy     241 PV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKK 294
               :     |  : |    |       | :   |  : | ||    | : | : | ||
Db     230 PVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIIASPVTANAIQQFVKQE 289

Qy     295 Y---NDGLAASAAPVFLALVGDTDVISGE--KGKKTKVTDLIYSAVDGDYFPEMYTFR 348
             |            :||| || |::||  :  || ::  :  : || :|  |: | |
Db     290 YEKEGND------------LTYVLLVGDHKDIPAKITPGIKSDQV--YGQIVGNDHYNEVFIGR 339

Qy     349 MSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKY-GMQYY 407
             :| :  |:    :    |   :: : | ||  :| :  || :|: |::|  :  ||:||
Db     340 FSCESKEDIKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQHENVIAN 399

Qy     408 YNQEHGYTDVYNYLKAPYTGCY-------------SHINTGVSFANYTAHGSETAWADPLLTT 457
             |  : ||  | :: :  :                 :||| :|  :  || | | |
Db     400 LLTQYGYTKIIK------CYDPGVTPKNIIDAFPNGGISLVNYTGHGSETAWGTSHFGT 451

Qy     458 SQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV---KEKGAYAYIGSSPNSW 513
             : :|: :|:  :     :: :|   : :   |     |     : || | :|: :|:
Db     452 THVQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDKPTGTVAIIASTINQSW 511

Qy     514 GEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNI--- 570
             |:| |:  |:|| ||: |                 |  | ::|   | | | : ||
Db     512 --------ASPMRG------------QDEMNEI------LCEKHPNNIKRT 536

Qy     571 -GNITHIGAHYYWEAYH---------VLGDGSVMPYRAMPKTNTYTLLPASLPQNQASY 618
              |:::| | || |  |            |||  :| |      |||  :||:|
Db     537 FGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQINLTDASV 596
```

```
Qy     619 SIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIKQ 677
              :||:  ||| | |  || :| |   ||| | |    ||| ||| || || :|| |:|
Db     597 NVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGVNKETVIKT 654

Qy     678 IQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK--KABGSREVKRIGDGLFVTIEPA 733
             |:|  | || ||||||:|||||||||||| ||:|     : :|  :|: | ::|  |
Db     655 INTNGENPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDA 714

Qy     734 ND-VRANEAKVVLAADNVWGDNTGYFLLDADHNTFGSVIPA-TGPLFTG-TASSNLYSA 790
              :   : ::  |   | :||:|| ||:||||: ||:|:||     | |   |::|:|:|
Db     715 PELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQVGQVIPSDTHTLWPNCSVPANLF-A 773

Qy     791 NFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVGYDYCITNPEPASGKMWIAGDGGNQPAR 850
             | |::|| |:|  |:|:|| |||:||| ||| ||||| |||| | :: | |||  |: ||
Db     774 PFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQG---PTK 829

Qy     851 YDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEED 910
             |||: | |||||::    |:|  || : | |:||:|| ||:|:|| ||||::: | || :
Db     830 EDDYVFEAGKKYHFLMKKMGGSGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTETTYRDA 889

Qy     911 GVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTL 962
             |:::| ::||||||: | | ||||||| ||| | || || | |:| |:||:|
Db     890 GMSAQSHEYCVEVKYAAGVSPKVCVDVYIPDGVADVTAQKPYTLTVVGKTITV 941
```

```
RESULT 5
CPG2_PORGI
ID   CPG2_PORGI     STANDARD;      PRT;    736 AA.
AC   P95493;
DT   15-DEC-1998 (Rel. 37, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Gingipain R2 precursor (EC 3.4.22.37) (Gingipain 2) (Arg-gingipain)
DE   (RGP-2).
GN   RGPB OR RGP2 OR PRTRII OR PG0506.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 230-651.
RC   STRAIN=HG66;
RX   MEDLINE=98370998; PubMed=9705298;
RA   Potempa J., Mikolajczyk-Pawlinska J., Brassell D., Nelson D.;
RA   Thoegersen I.B., Enghild J.J., Travis J.;
RT   "Comparative properties of two cysteine proteinases (gingipains R),
RT   the products of two related but individual genes of Porphyromonas
RT   gingivalis.";
RL   J. Biol. Chem. 273:21648-21657(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=22829867; PubMed=12949112;
RA   Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA   Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
RA   Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
RA   Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA   Dewhirst F.E., Fraser C.M.;
RT   "Complete genome sequence of the oral pathogenic bacterium
RT   Porphyromonas gingivalis strain W83.";
RL   J. Bacteriol. 185:5591-5601(2003).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX   MEDLINE=99452753; PubMed=10523290;
RA   Eichinger A., Beisel H.-G., Jacob U., Huber R., Medrano F.-J.,
RA   Banbula A., Potempa J., Travis J., Bode W.;
RT   "Crystal structure of gingipain R: an Arg-specific bacterial cysteine
RT   proteinase with a caspase-like fold.";
RL   EMBO J. 18:5453-5462(1999).
CC   -!- FUNCTION: Thiol protease which is believed to participate in
CC       intracellular degradation and turnover of proteins. Its
CC       proteolytic activity is a major factor in both periodontal tissue
```

```
CC       destruction and in bacterial host defense mechanisms. Activates
CC       complement C3 and C5 (By similarity).
CC  -!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
CC       immunoglobulins, with a preference for Arg in P1, and hydrophobic
CC       residues in P2 and P3.
CC  -!- SIMILARITY: Belongs to peptidase family C25.
CC       --------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       --------------------------------------------------------------------------
DR   EMBL; U85038; AAB41892.1; -.
DR   EMBL; AE017173; AAQ65700.1; -.
DR   PDB; 1CVR; 01-MAR-00; -.
DR   MEROPS; C25.003; -.
DR   TIGR; PG0506; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
KW   Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen;
KW   3D-structure; Complete proteome.
FT   SIGNAL        1     24
FT   PROPEP       25    229       BY SIMILARITY.
FT   CHAIN       230    736       GINGIPAIN R2.
FT   ACT_SITE    440    440       PROTON DONOR.
FT   ACT_SITE    473    473       NUCLEOPHILE.
FT   CONFLICT     58     58       G -> D (IN REF. 1).
FT   CONFLICT    246    246       P -> A (IN REF. 1).
FT   CONFLICT    251    251       E -> G (IN REF. 1).
FT   CONFLICT    254    254       E -> K (IN REF. 1).
FT   CONFLICT    398    398       I -> V (IN REF. 1).
FT   CONFLICT    435    435       A -> V (IN REF. 1).
FT   CONFLICT    480    482       YNV -> FSM (IN REF. 1).
FT   CONFLICT    510    510       N -> D (IN REF. 1).
FT   CONFLICT    512    512       S -> Y (IN REF. 1).
FT   CONFLICT    515    515       S -> N (IN REF. 1).
FT   CONFLICT    560    560       K -> N (IN REF. 1).
FT   CONFLICT    582    582       K -> E (IN REF. 1).
FT   TURN        237    238
FT   STRAND      240    245
FT   HELIX       247    252
FT   HELIX       254    262
FT   TURN        263    264
FT   STRAND      266    271
FT   HELIX       272    275
FT   HELIX       281    293
FT   TURN        295    296
FT   STRAND      299    305
FT   TURN        308    310
FT   STRAND      314    316
FT   TURN        317    318
FT   STRAND      319    321
FT   HELIX       323    326
FT   STRAND      338    342
FT   HELIX       347    362
FT   TURN        366    367
FT   TURN        369    372
FT   STRAND      373    378
FT   STRAND      382    385
FT   HELIX       387    389
FT   STRAND      391    391
FT   HELIX       392    406
FT   TURN        407    407
FT   STRAND      410    415
FT   HELIX       421    430
FT   STRAND      434    438
FT   STRAND      441    441
FT   STRAND      445    446
FT   TURN        447    449
FT   STRAND      452    452
FT   TURN        453    455
FT   HELIX       456    458
FT   TURN        462    463
FT   STRAND      467    471
FT   TURN        475    476
FT   TURN        478    479
FT   HELIX       485    491
FT   STRAND      493    494
FT   TURN        495    496
FT   STRAND      497    498
FT   STRAND      503    507
FT   HELIX       515    527
FT   TURN        528    529
FT   TURN        532    533
FT   STRAND      538    538
FT   HELIX       539    564
FT   STRAND      565    568
FT   TURN        571    572
FT   STRAND      574    574
FT   STRAND      574    581
FT   STRAND      585    586
FT   STRAND      590    591
FT   STRAND      593    594
FT   TURN        597    602
FT   STRAND      605    606
FT   TURN        608    613
FT   TURN        614    615
FT   STRAND      616    622
FT   STRAND      624    624
FT   TURN        625    626
FT   STRAND      627    631
FT   TURN        637    638
FT   STRAND      641    647
FT   TURN        649    650
FT   STRAND      651    651
FT   STRAND      654    660
SQ   SEQUENCE    736 AA;  80966 MW;  C848D3FAB420833 CRC64;

Query Match             5.9%;  Score 541;  DB 1;  Length 736;
Best Local Similarity  25.1%;  Pred. No. 3.5e-22;
Matches  200;  Conservative  130;  Mismatches  328;  Indels  140;  Gaps  29;

Qy      3  KLLLLIA-ASLLGVGLYAQSAKIKLDAPTTRITCTNNSFKQFDASFSFNEVELTKVETKG   61
                                                                        
Db      7  RIVSIVAFSSLLGGMAFAQPAE-RGRNPQVRLLSAEQSMSK--VQFRMDNLQFTGVQTSK   63

Qy     62  GTFASVSIPGAFPTGEVGSPEYPAVRKLIAVPGATPVVRVKS--FTEQVYSLNQYGSEK  119
                                                                        
Db     64  GVAQVPFTEGVNISEKGTPILPILSRSLAVSETRAMKVEVVSSKPIEKKDVL------   116

Qy    120  LMPHQPSMSKSDDPEKVPFVINAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQY  179
                                                                        
Db    117  IAPSKGVISRAENPDQIPYVVGQSYNEDKFFPGEIATLSDPF-ILRDVRGQVVNFAPLQY  175

Qy    180  DVVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYN  239
                                                                        
Db    176  NPVTKTLRIYTEIVVAVSETAEAGQNTISLVKNSTFTGFEDIYKSVFMNYEATRY-----  230

Qy    240  TPV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHK  293
                                                                        
Db    231  TPVEEKENGRMIVIVPKKYEEDIEDFVDWKNQRGLRTEVKVAEDIASPVTANAIQQFVKQ  290

Qy    294  KY----NDGLAASAAPVFLALVGDTDVISGE-KGKKTKKVTDLYYSAVDGDYFPEMYTF  347
                                                                        
Db    291  EYEKEGND------LTYVLLVGDHKDIPAKITPGIKSDQV---YGQIVGNDHYNEVFIG  340

Qy    348  RMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKV-GMQY  406
                                                                        
Db    341  RFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQHENIIA  400
```

```
FT   NON_TER          1
FT   DOMAIN           1      39     HYR 1.
FT   DOMAIN          40     123     HYR 2.
FT   DOMAIN         124     207     HYR 3.
FT   DOMAIN         208     292     HYR 4.
FT   DOMAIN         293     376     HYR 5.
FT   DOMAIN         377     460     HYR 6.
FT   DOMAIN         461     544     HYR 7.
FT   DOMAIN         546     629     HYR 8.
FT   DOMAIN         630     713     HYR 9.
FT   DOMAIN         714     797     HYR 10.
FT   DOMAIN         798     881     HYR 11.
FT   DOMAIN         882     966     HYR 12.
FT   DOMAIN         967    1050     HYR 13.
FT   DOMAIN        1051    1133     HYR 14.
FT   DOMAIN        1134    >1200    HYR 15.
FT   NON_TER       1200
SQ   SEQUENCE    1200 AA;  124742 MW;  57A52B7CDE68A629 CRC64;
```

```
Query Match            2.7%;   Score 250.5; DB 1;  Length 1200;
Best Local Similarity  20.3%;  Pred. No. 6.5e-06;
Matches  285;  Conservative 146;  Mismatches 499;  Indels 471;  Gaps   69;
```

```
Qy    430  SHLNTGVSFANYTAHGSETAWADPLLTSQLKALTNKDKYFLAIGNCCIT---AQFDVVQ  486
                ||||||       |||         |: : :|| ::||         :| |
Db      2  SH-NPGQSFTTGT--------------------TTTVVYTATDA--FANVGQCAFTITVTATDTTP   44

Qy    487  PCF---GEVITRVKEKGAYAYIGSSPNSYWGEDYYWSGANAVFGVQPTFEGTSMGSYDA  543
                |||           :||  ||||       |||
Db     45  PVVTVSSSTITRQVEVG-----------------------TPGVN-VFFTEPTATDNS----GQ   80

Qy    544  TFLEDSYNTVNSIMWAGNLAATH-----AGN--IGNITHIGAHYWEAYHVLGDGSVMPY  596
                |||                              :||
Db     81  AILISRTNQPGDFFSVGQTIVTYTFQDPSGNPASGTVT----------INVIEVDTTPPV  130

Qy    597  RAMPKINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQI  656
                ||       ||||    ||:||        |  :   |
Db    131  VTVPST------------------------TITRQVEVGS------------------SGVNVFYTEPTA  158

Qy    657  TENGNYDVVITRSNYLPVIKQIVQGEPSPYQPVSNLTATTQQKVTLKWEAPSAKKAEGS  716
                |||||:|  :|    ::||         |||
Db    159  TDNSGQAILTSRTN---------------------QPGDFFPVGQT-----------VTYTFQDPSGNPASGT  200

Qy    717  REVKRIGDGL---FVTIEPANDVRANEAKVLAADNVW-----GDNTGYQFLLDADHNT  767
                |||                            |||
Db    201  VTINVIEVDTTPPVVTV-PSTTI-TRQVEVGSSGVNVFYTEPTATDNSG-QAILTSRTNQ  257

Qy    768  FGSVIPATGPLFTGTASSNLYSANFEYLVPANA---------DPVVTQNIIVTGQG  815
                ||||                 |||               |||
Db    258  PGDFFPVGQTVTYTFSQD-PSGN----------PASATVTINVIEVDTTPPVVTVPSTTITRQV  311

Qy    816  EVVIPGGVYDYCITNPEPASGKMWIAG--DGGNQPARYDDFTFEAGK--KYTFTMRRAG  870
                ||:|            ||:||              ||
Db    312  EIGTPGVNVFY--TEP-TATDNSGIANLVSRTNQPG---DF FPVGQTDVYTFQDPSGN  364

Qy    871  MGDGT----DMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT  926
                ||                     |||
Db    365  PSSGTVTINVEVDTTSP---VITVPSN-------------------TITREVELGTPG------VNVFYT  407

Qy    927  AGVSPKVCKDVTVEGSNEFAPVQNLTGS-------SVGQK-VTLKWDAPNGTPNPNPNPN  980
                |||                     :: :|
Db    408  ------EPTATDNSGIANLVSRTNQPGDFFNVGQTVVTYTFQDPSG-------NPA  450

Qy    981  PGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLT  1040
                :| |                              |||
Db    451  SGT-----------VTINVEVD-----------------TTPPQVFVTGEVVRTAPFGSSGV----  486

Qy   1041  PDNYLITPALDLPNGGKLITFWVCAQDANYASEHYAV------YASSTGNDASNFTNAL  1092
                 |||                                        |||
Db    487  --NVFFTEPVAIDNSG--TADLVSQTAQ-PGDFFPIGSTVVTYYRDASGNEATGTFTVT  541

Qy   1093  LEE------TITAKG---VRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY  1142
                :||             :   :    | :     |::
```

```
Qy    407  YYNQEHGYTDVYNYLKAPYTGCY---------SHLNTGVSFANYTAHGSETAWADPLLT  456
                :: ||:
Db    401  NLLTQYGYTKIIK--------CYDPGVTPKNIIDAFNGGISLANYTGHGSETAWGTSHFG  452

Qy    457  TSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV---KEKGAYAYIGSSPNSY  512
                  :|
Db    453  TTHVKQITNSNQLPFIFDVACVNGDFLYNVPCFAEALMRAQKDGKPTGTVAIIASTINQS  512

Qy    513  WGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNI--  570
                |
Db    513  W-------ASPMRG---------------------------LCEKHPNNIKR  537

Qy    571  --GNITHIGAHYYWEAYH---------VLGDGSVMPYRAMPKINTYTLPASLPQNQAS  617
                  ||:        ||   :: :
Db    538  TFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFRGDPSLLVRTLVPTKMQVTAPANISASAQT  597

Qy    618  YSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIK  676
                || :||     : :|||                |:|
Db    598  FEVACDYNGAIATLSDDGDMVGTAIVK-DGKAIIKLNESIADETNLTLTVVGYNKVTVIK  656

Qy    677  QIQVGEPSPYQPVSN---LTATTQGGKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPA  733
                |:    ::      ||            |:|
Db    657  DVKV-EGTSIADVANDKPYTVAVSGKTITV--ESPAA-----------------GL-TIFDM  697

Qy    734  NDVRANEAK--VVLAADN  749
                ||
Db    698  NGRRVATAKNRMVFEAQN  715
```

```
RESULT 6
HYAL_STRPU
ID   HYAL_STRPU     STANDARD;     PRT;   1200 AA.
AC   O76536;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hyalin (Fragment).
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98142007; PubMed=9473317;
RA   Wessel G.M., Berg L., Adelson D.L., Cannon G., McClay D.R.;
RT   "A molecular analysis of hyalin -- a substrate for cell adhesion in
RT   the hyaline layer of the sea urchin embryo.";
RL   Dev. Biol. 193:115-126(1998).
CC   -!- FUNCTION: Major constituent of the hyaline layer. The hyaline
CC       layer of echinoderm embryos is an extraembryonic matrix that
CC       functions as a substrate for cell adhesion through early
CC       development.
CC   -!- SUBUNIT: Homooligomer in presence of calcium.
CC   -!- SUBCELLULAR LOCATION: Secreted. Extracellular matrix.
CC   -!- PTM: Glycosylated.
CC   -!- SIMILARITY: Contains at least 15 HYR domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF076472; AAC31909.1; -.
DR   PIR; T17404; T17404.
DR   InterPro; IPR003410; Hyalin.
DR   Pfam; PF02494; HYR; 15.
DR   PROSITE; PS50825; HYR; 15.
KW   Cell adhesion; Extracellular matrix; Glycoprotein; Repeat.
```

Db   542 VEEMVDTTPPTVTVSGGNIVRTVELGQSRLSVIYTEPTATDNSGEANLVSRTAQPGDLFP 601

Qy   1143 IDLDEVEI-----KANGKRADFTET----------------------FESSTHGEAPA-----EW 1175
                : | | :   : |:|::|
Db   602 VGTTTVEVYVQDPAGNEGRGQFTVTVIAVDTTPPSVNCPNNVFQEVELGTSSAPVFFTEP 661

Qy   1176 TTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVK 1235
                |:|:| | :|: |: :|
Db   662 TAFDISGQ-----------------------ANLVTRTAAPG------DSFPV-----GTTSVS 691

Qy   1236 YYYAVNDGFPGD--HYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQ 1293
                | || |:| | ||| :| | :|| :| | |
Db   692 YIFSDNSGNEAEPCTFTVTISAVDTT-------PPTVNCINN------------------------ 726

Qy   1294 SVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIK 1353
                : :|||:| | |:
Db   727 ---VAQTVQLGTGSTQVFFTEPTAFDIS----------------------GQTSLVTRTSAPGDSFP 768

Qy   1354 EGLTETTPEEDGVATGNHEYCVEVKYTAGVSPKKCVDTVNSTQFNP-----VQNLTAEQ 1408
                :| ||| | :| :| |:
Db   769 VGTTSVYIFTDASGNNAQPCT-------------------------FNVVINAVDTTPTVNCANNIAQTI 815

Qy   1409 APNSMDAILKWNAPASKRAEVLNEDFENGIPASWKTIDADGDGNNWTTTPPPGGSSFAGH 1468
                :|: :|: |: || |:| | | |::
Db   816 ELGSTSAVVNYAEPSA-------------------------TDISGTAFLVSRSSSPCDSFPIGS 855

Qy   1469 NSAICVSSASHINFEGP---------------------QNPDNYLVTPELSLPGGGTLTFWVC 1510
                | || || || | |: |:| |||||:|:|||:|
Db   856 TTVTYIFSDQSGNEAPPCVFTVTIGTVDTMMPFTSCPNNIVQSVELGVP--GTVISW-- 911

Qy   1511 AQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQCTWYQKTV 1570
                | : | :   ::| :| |:|
Db   912 -----------------TTPTANDAAGIA--------------------------SIVSNLQPGSFF--TV 937

Qy   1571 QIPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGV-TETTYRD 1629
                |:|||| | ||:   : | |: | | | ::||::
Db   938 GESATVTVATDNSGLTD-----------------------NSCSFTVTFAVDTTPPSVVCTNNVFQT 983

Qy   1630 PDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTIITVTCQGEAM 1689
                |: || |:|:| | | | : | ::| |: |: || :| |
Db   984 VELGTN---FVQVFYTEPTASDISGQANLVSRTNV----PGDSFPVG-TSTVT-----YV 1030

Qy   1690 IYDMNGRRLAAGRNTVVYTAQ 1710
                |: :| | ||:|::|:|:|||
Db   1031 FADNSGNTADACSFTIVTAQ 1051

RESULT 7
WAPA_BACSU
ID   WAPA_BACSU   STANDARD;   PRT;   2334 AA.
AC   Q07833;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Wall-associated protein precursor.
GN   WAPA OR N17G OR BSU39230.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=93302506; PubMed=8316082;
RA   Foster S.J.;
RT   "Molecular analysis of three major wall-associated proteins of
RT   Bacillus subtilis 168: evidence for processing of the product of a
RT   gene encoding a 258 kDa precursor two-domain ligand-binding
RT   protein.";
RL   Mol. Microbiol. 8:299-310(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE=95219088; PubMed=7704263;

RA   Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT   "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT   genome containing the hut and wapA loci.";
RL   Microbiology 141:337-343(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE=97124196; PubMed=8969509;
RA   Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA   Miwa Y., Fujita Y.;
RT   "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT   containing the lic and cel loci, and creation of a 177 kb contig
RT   covering the gnt-sacXY region.";
RL   Microbiology 142:3113-3123(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi Y., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC   MOTILITY, SECRETION OR DIFFERENTIATION.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC   INTO THE MEDIUM.
CC   -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC   101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC   C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC   MOTIF REPEATED 31 TIMES.
CC   -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC   SIMILARITY TO THE REPEAT IN E.COLI RHS GROUP OF PROTEINS (RHSA-D).
CC   --
CC   --
DR   EMBL; L05634; AAA22883.1; -.
DR   EMBL; D31856; BAA06656.1; -.
DR   EMBL; D29985; BAA06260.1; -.
DR   EMBL; D83026; BAA11683.1; -.
DR   EMBL; Z99124; CAB15959.1; -.

```
DR   PIR; S32920; S32920.
DR   SubtiList; BG10797; wapA.
DR   InterPro; IPR003305; CBM_CenC.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF02018; CBM_4_9; 1.
DR   Pfam; PF05593; RHS_repeat; 14.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 17.
KW   Cell wall; Repeat; Signal; Complete proteome.
FT   SIGNAL         1     28
FT   CHAIN         29   2334       WALL-ASSOCIATED PROTEIN.
FT   DOMAIN       504    869       3 X 101 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT       504    605       1-1.
FT   REPEAT       636    736       1-2.
FT   REPEAT       769    869       1-3.
FT   DOMAIN      1021   2139       31 X 21 AA APPROXIMATE TANDEM REPEATS OF
FT                                 X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
FT   REPEAT      1021   1040       2-1.
FT   REPEAT      1042   1061       2-2.
FT   REPEAT      1063   1082       2-3.
FT   REPEAT      1083   1102       2-4.
FT   REPEAT      1109   1128       2-5.
FT   REPEAT      1129   1148       2-6.
FT   REPEAT      1150   1169       2-7.
FT   REPEAT      1174   1193       2-8.
FT   REPEAT      1199   1218       2-9.
FT   REPEAT      1219   1238       2-10.
FT   REPEAT      1646   1665       2-11.
FT   REPEAT      1667   1686       2-12.
FT   REPEAT      1690   1709       2-13.
FT   REPEAT      1711   1730       2-14.
FT   REPEAT      1732   1751       2-15.
FT   REPEAT      1753   1772       2-16.
FT   REPEAT      1795   1814       2-17.
FT   REPEAT      1820   1839       2-18.
FT   REPEAT      1840   1859       2-19.
FT   REPEAT      1861   1880       2-20.
FT   REPEAT      1887   1906       2-21.
FT   REPEAT      1908   1927       2-22.
FT   REPEAT      1929   1948       2-23.
FT   REPEAT      1969   1982       2-24 (APPROXIMATE).
FT   REPEAT      1983   2002       2-25.
FT   REPEAT      2008   2027       2-26.
FT   REPEAT      2028   2047       2-27.
FT   REPEAT      2051   2070       2-28.
FT   REPEAT      2071   2090       2-29.
FT   REPEAT      2093   2112       2-30.
FT   REPEAT      2120   2139       2-31.
SQ   SEQUENCE   2334 AA;  258329 MW;  B75138CCD278BAA3 CRC64;

Query Match             2.6%;   Score 242.5;  DB 1;  Length 2334;
Best Local Similarity  19.4%;   Pred. No. 4.9e-05;
Matches  424;  Conservative 265;  Mismatches 750;  Indels 743;  Gaps 118;
```

```
Qy   14  GVGLYAQSAKI----------------------------KLDAPTRTTCTNNSFKQFD---------A   45
         ||| |::|                               |:
Db   88  GDGVYKQEVYLDPIHTKETPDADWEDISPELKESTSKQVETENAILNSDFQKQMKNGLYA         147

Qy   46  SFSPNEVELT--KVETKGGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVK   103
         :|  : :|:   |:
Db  148  TFFHNDHKVTYSLAEAKGPNKTSLITPKDTSADYKTDSNEI------VYPDVFPNIDLQ    199

Qy  104  SFTEQVISLNQYGSEKLMPHQ--------------PSMSKSDDPEKV----------    136
         :|:  :   : ||:
Db  200  TFT-----FNENIKEDLVLHQYNGYNTFTFQLKTDLQAKEQEDGSIDFSDEKGKVVFSVP   254

Qy  137  -PFVYN------AAAYARKGFVGQELTCVEMLGTMRGV-----RIAALTINPVQ       178
          ||:                   : | :     :
Db  255  KPFMTDSKLDELSGEVERSDKIVSYKLEKNEE-GYLLHITADENWLKDPERVVPVSIDP--   311

Qy  179  YDVVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPY-FETAYKQLFNRDVYTDHGDL   237
         :|  ::         : :                     |:|  : :       |  |:
Db  312  --STSLSVSSDTFVMSAYPFTNYSASSQKWDANLKAYVLKTGY---YDKTGTNYAFM      364
```

```
Qy  238  -YNT--PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDV---HYTDEAEV--GTTNASI-   287
          ::|:  ||  |   |:||  :| | |       |  |::|  |  | |:|
Db  365  KFNNLKPIQNMVTKATLKTVAHSYGTKATGLWLDTVNSNY-DNAKVTWNTKPASKNI      423

Qy  288  -KAFIHKKYNDGLAASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYT   346
          || |||                                    :::   :   |
Db  424  GKADVHK-----GQWAS------------------YDVTAAVKS-----WNSGGANYGFKLHT  458

Qy  347  FRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQY   406
         :  :                                                :  | |
Db  459  ---NGNGKEYWKKLI------SSANSANKPYIEVTYTIPKGN------TPTIK--AY       498

Qy  407  YYNQEHGYTDV--------YNYLKAPTGC-YSHLNTG------VSFA               439
          :                  | | | |
Db  499  HNGDSTGYFDISWKKVEGAKGYKVWIYNGKEYQAISAGNVTSWSTKGKKIWPTSAEFASK   558

Qy  440  NYTAH------GSETAW-ADPLLTTSQLKALTNKDKYPLAI------GNCCITAQFDVQP   487
         :  |:        |   |:|::|:|                           : | |  |
Db  559  RYKLHLDGKDGAELALDPSPVYKNSGGSYATSKN-YWIGVSAIFDQGEGAMSAPAKPVIP   617

Qy  488  CFGEVITRVKEKGAYAYIGSSPNSYWGEDYWSVGANAV-FGVQ-----PTFEGTSMGSYD   542
                 :  :|  |   :  :|                 | |:
Db  618  NVG----KAQAPSAKGYNNGNATGYP--DLSWKAVSGATGYKVQVFNGKGFETLDLGN--   669

Qy  543  ATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVL-GDGSVMPYRAMPK   601
                   :         |      |     |:             |:|:
Db  670  ----QTSWTTKGKKIWPTS-AEIKAGKY------ALHLKDGSGAELPINPGP-          710

Qy  602  TNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVAT------VSMTKQI  656
          |::              ::      |          |  |        |
Db  711  --TYK-NAGGDGAKRNYSFK-------IIAYNKD-----GEAIASPAATPALPDIARPKNV  756

Qy  657  TENGNYDVVITRSNYLPVI--------KQIQVGEPSPYQPVSNLTA--TTQGQKVTLK     704
          : ||||           :        | : |:        :|         |  |:
Db  757  TGYLYTNTKSSQTGYVNLIWEKVQNAKGYKVNIYNGKEYQSFDVGDADHWTTQNKNI---  813

Qy  705  WEAPSAKKAEGSREVKRIGDGLFVTIEPA---NDVRANEAKVVLAADNVWGDNTGYQFLL   761
         | :| : | |   :        :  | |:       |: | :  |    |:|  ||:| |
Db  814  WPTSEEIKA-GSYKLHTDGKGGELALIDPSPVYNNANGN------YKGKKNYSFTL       861

Qy  762  DADRHTFGSVIPAT------GPLFTGT--------ASSNLYSANFEVIVP-            797
         :|                  |::|| |           ::
Db  862  VA--YDANGETIPTAPFNPTFHEGAEFLGTEBYWSIIDIPSGQLNGATGNVIVNEEDLSID  920

Db  798  ----------------------------ANADPVV-----------------TQNIIVTGQGE  816

Db  921  GRGPGLGLSRTYNSLDSDSDHLFGGQWYADAETSVISTDGGAMYIDEDATTHRFTKKADGT   980

Qy  817  VVIPGGVY-------------DYCI----TNP-EPASGKMWIAGDGGNQPA--RYDD----   853
         |:|||||              :|:        ::|
Db  981  YQPPTGVYLELTETADQFLIKTDQTNAYFNKKGGKLQKVVDGHNNATVVTYNDKNQLTA   1040

Qy  854  FTFEAGKKYTFT--------------------------MRRAGMGDGTDMEVEDDS----   883
          : ||  |:||
Db 1041  ITDASGRKLTFTYDENGHVTSITGPKNKKVTYSYENDLLKKVTDTDGTVTSYDYDSEGRL  1100

Qy  884  ---------PASYTYTYVR-DGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKV   933
                  |:|| | :  ||| :|| |        | :|     |   : :
Db 1101  VKQYSANSTEAKPVFTEYQYSGHRLEKAINAK--KETYV----YSYDADKKTLLMTQPNG  1154

Qy  934  CKDVTVEGSNEFA-PVQNLTGSSVGQKVT--LKWDAPNGTPNPNPNPNGT-TLSESF      989
                 | :| | |:  :::   :|:|:   | ::| |      :       |:|
Db 1155  RK--VQYGYNEAGNPIQ-VIDDABGLKITTNTKYBGNNVVEDVVDPN---DVGTGKATESY 1208

Qy  990  ENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPA  1049
         :|  |  |  ||:|||  | :|     ||  :  :|:| :       |::| |
Db 1209  Q----------YDKDGNVTSVKDAYGTETYEYNKNNDVTKMKDTEGNVTDIAYD------G 1253

Qy 1050  LDLPNGGKLTFWVCAQDANYASEHYAVYAS-----STGNDASNFTNALLEETTAK----  1100
         || ||| |   :        :   |:|| :     |||| :| :  ||: ::||
Db 1254  LDA----------VSETDQSGKSSSAAVDKYGNQIQSSKDLSASTNILKDGSFEAQKSGW  1304

Qy 1101  --------------GVRS-PKAIRGRIQGT--------------WRQKTVDLPAGTKYVAF 1132
```

```
Db  1305 NLTASKDRRKISVIADKSGVLSGSKALEVLSQSTSAGTDHGYSSATQTVELEPNTTYTLS 1364
              ||:|                   |:       |
Qy  1133 RHFQSTDM------FYIDLDEV----------EIKANGKRADFTE---TFESSTHGEAPA 1173
              :|:|          ||:.|
Db  1365 GKIK-TDLAKSRAYFNIDLRDKDQKRIQWIHNEYSALAGKNDWTKRQITFTTPANAGKAV 1423

Qy  1174 EWTTI---DADGDGQGW--LCLSSGQL--DWLTAHGGSNVVSSFSWNGMALNPDN----- 1221
              ::|                        :
Db  1424 VIMEVDHKDKDGDKGKANWFDEVQLEKGEVSSSYNPVQNSSFTSATENWNVSGASVDSEEGF 1483

Qy  1222 ------------------------YLISKDVTGATKVKY----------------- 1236
                                      ||||:
Db  1484 NDDVSLKAARTSASQAGSVTKQTVLQGSANDKPVYLTLTGMSKASSVKFTDEKDYSLQA 1543

Qy  1237 --YYA-----VNDGPFG-----DHYAVMISKT-GTNAGDFTVFEETPNGINKGGARFG 1282
              :||||  :: ::|||:|
Db  1544 NVTYADGSTGIYNAKFPSGTQEWNRAAVVIPKTKPINKVDISILFQKSATG------- 1594

Qy  1283 LSTEANGAKPQSVWIE--RTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPFPTD 1340
                                 || :::|
Db  1595 -----TVWFDDIRLIEGSLLITKST---YD-SNGNYVTKEEDEL--GYATSTD 1635

Qy  1341 Y---------TYT-------VYRDGTKI-----KEG-LTETTFEEDGVA 1367
              |           :|       |:
Db  1636 YDETGKKTSETDAKGEKTTYTYDQADQLTNMTLSNGTSILHSYDKEGNEVSKTIRAGADQ 1695

Qy  1368 TGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRA 1427
              :|    ||||  :::|    ||  :::
Db  1696 TYKFEYDVMGKLVKTTDPLG---NVLASEYDANSNLTKTISPNGNEVSLSYDG------T 1746

Qy  1428 EVLNEDFENGIPASWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVSSASHINFEGPQN 1487
              :||                                   :
Db  1747 DRVKSKSYNGTEKYIFTYDKNG----------NETSVVNKE-QN 1779

Qy  1488 P----------DNYLVTPELSLPGGGLTFWCAQDANYASEHYAVYASSTGNDASNFANALLE 1541
              :          |                  :     :
Db  1780 TTKKRTFDNKNRLTELITDRGGSQT--WTYPSDSDKLKTFSWIHGDQKG--TNQFTYNKLD 1835

Qy  1542 EVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLP---------VAFRHFGCT 1587
              :::                              :
Db  1836 QMIEMKD-------STSSYSFDYDENGNVQTFITCNGGGTSFSYDERNLVSSLLHIGDK 1886

Qy  1588 DFFWINLDDVVITSGNAPSYYTIYRNNTQI--ASGVTETTYRDPDLATGFYTYG---- 1640
              |                                            :
Db  1887 -----NGGDIL-----TESYEYDANGNRTTINSSASGKVQ-----------YEYGKLNQ 1924

Qy  1641 -VKVVVPNGESAIETATLNITSLADVTAQKPYTLTVVGKTIIVTCQ-----GEAMIYDMN 1694
               |            ::|                  :::|
Db  1925 LVKETHEDG-TVIEYTYDGFGNRKTVTTIKDGSSSKTVNASFNIMNQLTKVNDESISYDKN 1983

Qy  1695 GRRLAAGRNTVVYTAQGGHYAV 1716
               ::|                ||
Db  1984 GNRTSDGKFTYTWDAEDNLTAV 2005
```

```
RESULT 8
YD96_METJA
ID   YD96_METJA    STANDARD;    PRT;   2894 AA.
AC   Q58791;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hypothetical protein MJ1396.
GN   MJ1396.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=9637999;
```

```
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
```

```
DR   EMBL; U67579; AAB99406.1; -.
DR   PIR; C64474; C64474.
DR   TIGR; MJ1396; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001220; Lectin_legB.
DR   InterPro; IPR006626; PbH1.
DR   Pfam; PF00139; lectin_legB; 1.
DR   SMART; SM00710; PbH1; 20.
KW   Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT   TRANSMEM      8     28       POTENTIAL.
FT   REPEAT      543    567       PBH1 1.
FT   REPEAT     2085   2107       PBH1 2.
FT   REPEAT     2135   2156       PBH1 3.
FT   REPEAT     2158   2180       PBH1 4.
FT   REPEAT     2201   2223       PBH1 5.
FT   REPEAT     2224   2244       PBH1 6.
FT   REPEAT     2245   2266       PBH1 7.
FT   REPEAT     2267   2289       PBH1 8.
FT   REPEAT     2290   2311       PBH1 9.
FT   REPEAT     2341   2363       PBH1 10.
FT   REPEAT     2367   2389       PBH1 11.
FT   REPEAT     2390   2419       PBH1 12.
FT   REPEAT     2422   2444       PBH1 13.
FT   REPEAT     2455   2477       PBH1 14.
FT   REPEAT     2479   2501       PBH1 15.
FT   REPEAT     2512   2542       PBH1 16.
FT   REPEAT     2550   2582       PBH1 17.
FT   REPEAT     2589   2611       PBH1 18.
FT   REPEAT     2612   2633       PBH1 19.
FT   REPEAT     2638   2660       PBH1 20.
SQ   SEQUENCE   2894 AA;  322361 MW;  8B64F78G537A16DF CRC64;

Query Match             2.4%;  Score 224;  DB 1;  Length 2894;
Best Local Similarity   18.2%;  Pred. No. 0.0007;
Matches  312;  Conservative  213;  Mismatches  562;  Indels  626;  Gaps  90;
```

```
Qy   214 FSPYFETAYKQLFNRDVYTDHGDLYNTPVRMLVVAG---AKFKEALKPWLTWKAQKGFYL 270
             ::||     |:          ::|||||:          |||:|:||
Db   792 WSTWSDIRYDNLNNDSSY-EGD------------------AGTALAWTKDSLKPGEIWVPIIWGL 838

Qy   271 DVHYTD---EABVGT---NASIKAFIHKKYNDGLAASAAPVF----LAlVGDTDVISG 319
             |:|         ::         :|       ||    :::|||||
Db   839 GYNYTDMNNEINMGLSQLYDTGVKSIDYPNNGDSFNPNIGPIIYINSTIALYGLVDAYNL 898

Qy   320 EKGKGKTKKVTDLYYSAVDGDYF----------PEMYTFRMSASS-PEELTNIIDK 363
               |            ::|                  :
Db   899 N---------VSINITQNGTYIYTNSTLINLSVPYEEKLVSFPVNISNMPYGAYNITIK 950

Qy   364 V-LMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVN--- 419
               :            :|:|                  :::|
Db   951 TNLPNDQNTSND----EKSIIIYITSFSVQPNYQEKTKGNVGBEIFYN-----ITLYNFGV 1001
```

```
Qy   420 ---------------YLKAPYT--------------GCYSHLNTGVSFAN------YTAHGS-  446
                                            :  :::|:|      :
Db  1002 GGRFDINITYLTKGWTKIYNNSILIAEDANGDGIWDYINPNYDLNSNNLPDIYVPTGEI 1061

Qy   447 ETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDY--------VQPCF---GEVIT  494
Db  1062 NLTVSKTIPSTAPLGEIDTTTLKFVNINNPSIFGKTFQTSTPYPPSVQKTFYLHGDTLR 1121

Qy   495 RVKEK----GAYAYIGSSPNSYWGE----DYYWSVG---------ANAVFGVQ-----  530
Db  1122 TLNTSIPTTINNYTTINSNSLASWICYPRFADNFTVVGKIPILLYINDPNVIFGTEMHKI 1181

Qy   531 ------PTFEGTSMGS-YDATFLEDS-----YN-TVNSI--------MWAGNLAATH  566
Db  1182 VVSLMATNGIIDSFTLGSDVEYLYLDDTIKSYIFNITLDSIITIPKNYILVLRVENQISSN 1241

Qy   567 AGNI------GNITHIGAHYWEATHVLGDGSVMEPYRAMPKTNYTLPASLPQNQASYS  619
Db  1242 SINIYHNSTYPSNIT-LNTTTYVNVYNIFSDKNVY----LPNENV-TIFANTDPIGSYD 1295

Qy   620 IQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG------NYDVIIRS-NY  671
Db  1296 I--SGANITVYYPNGSVY-------INSSMLLQEIDKNSPSLWKLYNYSFSLPESGKY 1344

Qy   672 LPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKABGSREVKRIGDGLFVTI-  730
Db  1345 LITITGIE----------SNGVISKKNYTIYCGYEIQGYVKED---------FGTLG 1382

Qy   731 -EPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGGSVIPATGPLFTGTASSNLYS  789
Db  1383 KEDSEDKGIYGVNVSLLEDS--------NNDGIPDIGD-------TIVNSTTTDIFGHYSFLVYN 1432

Qy   790 ANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVDYCITNPEPASGKMWIA-GDGGNQP  848
Db  1433 SSKTYFVVVNSRTVGTTRGLNPQYSKNDIWAEETYQTVYT---PINSSQWIANGNASIFP 1489

Qy   849 ARY---DDFTFEAGKKYTFMRRAGMGDGTDMEVE---------DDSPASYTYTVRDGT  896
Db  1490 DKLLLTTDDYG-EAGSVWYY--KPVNLSE-DLVVEFYAYLGDNPDGADGITFTLQSLGT 1544

Qy   897 KIKEGLTATTFEEDGVAAGNEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSV  956
Db  1545 -----NELGGTGGDLGY-------GGISPSVAVEVDT-WLNDF-------------- 1574

Qy   957 GQKVTLKWDAP---------NGTPNPN-----PNPNPNPGTTLSESFENGIPASW---  997
Db  1575 --------DAPATTDHIAIDVNGNINHTYNSLTYPTPNPYDLGNVEDGREHLIKIVWNAT 1626

Qy   998 -KTIDADGDGH---GWKPGNAPGIAGYNSNGCVYSESFGLGG---------IGVLTPDNY 1044
Db  1627 TKTLQVYFDGNLSLTWN----KDITQIIGNSAYFGFTGGTGGAKNLQYVKPIYVKNGDGY 1682

Qy  1045 LITPALDL--------PNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEET 1096
Db  1683 IINPTYGVVEMFGGRDPN--EEDNW---EDGKY-INLNSYSGKNITFGFSFDV 1734

Qy  1097 IT--------AKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1146
Db  1735 ITNTKSTGQGSFSQFIKNANAIYGKDESYFRIPNIDAKNGNHYI---YTSGNKI---LD 1787

Qy  1147 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNV 1206
Db  1788 NLTI-VNGS--------------------------TQINGTII 1803

Qy  1207 VSSFSW--NGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 1264
Db  1804 LSGLQWIANGNA-----YI-------NNSNNLT 1824

Qy  1265 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 1324
Db  1825 LIL-TPDDYNQKG--------SVW-------YYKPVNLSEDL 1850

Qy  1325 LDDIQFTMGGSPTPTD-YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383
Db  1851 VVEFYAYLGDNPDGADGITFTLQSLGT--------NELGGTGGDLGY---------GGI 1892

Qy  1384 SPKKCVDTVNSTQF--------------NPVQNLTAEQAPNSMDA---------IL 1417
Db  1893 SPSVAVEVDTWLNDFDSPATTDHIAIDVDGNLNHTYNSLTYSTPNPYDLGNVEDGREHLI 1952

Qy  1418 K--WNAPASKRAEVLNEDFENGIPASWKTIDADGDGNNWTTPPPGGSSFAGHNSAICVS 1475
Db  1953 KIVWNA---TKTLQVYFDGNLALTW-------NKDITQIIGNSTYFGFTGG--TG 1996

Qy  1476 SASHINFEGP--QNPDNYLVTPELS------------------LPGGG 1503
Db  1997 GAKNLQYVKPTYVKNGDNVLNLEEISPNPIIDNVGADTYIGNIFFENVSVGILGNETGLN 2056

Qy  1504 TLTFWVCA-------QDANYASEHYAVYASS-TGNDASNFANALLEEVLTAKTVV 1550
Db  2057 NLTLKSCGIYGKILNAGVKLVDYDWSLQNYPLYIDNLTINASGGYGISMLNKI------ 2109

Qy  1551 TAPEAIRGTRAQGTW--YQKTVQLPAGTKYVAFRHFGCTDFFWIN--------LDDVV 1598
Db  2110 ----------------WAMLYNSQISLKNG------------VGIYWANWAGGFGNITIYNIT 2144

Qy  1599 ITSGNAPSYTYT----TYRNNTQIASGVTETTY 1627
Db  2145 ISSCNQGLVLYKDGNGIKLINSQIKNSVYEGVY 2177
```

RESULT 9

```
IGA1_HAEIN
ID   IGA1_HAEIN     STANDARD;      PRT;   1541 AA.
AC   P42782;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN   IGA.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HK368 / Serotype B;
RX   MEDLINE=89379374; PubMed=2506130;
RA   Poulsen K., Brandt J., Hjorth J.P., Thoegersen H.C., Kilian M.;
RT   "Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
RT   of Haemophilus influenzae serotype b.";
RL   Infect. Immun. 57:3097-3105(1989).
RN   [2]
RP   MUTAGENESIS OF SER-288.
RC   STRAIN=HK368 / Serotype B;
RX   MEDLINE=92234949; PubMed=1373717;
RA   Poulsen K., Reinholdt J., Kilian M.;
RT   "A comparative genetic study of serologically distinct Haemophilus
RT   influenzae type 1 immunoglobulin A1 proteases.";
RL   J. Bacteriol. 174:2913-2921(1992).
CC   -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC       PRODUCING INTACT FC AND FAB FRAGMENTS.
CC   -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC       certain Pro-|-Xaa bonds in the hinge region. No small molecule
CC       substrates are known.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC       SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC       OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC       DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: Belongs to peptidase family S6.
```

```
DR   EMBL; X64357; CAA45708.1; -.
DR   EMBL; M87492; AAA24969.1; -.
DR   PIR; A37023; A37023.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   InterPro; IPR000710; Peptidase_S6.
DR   InterPro; IPR004899; Pertactin.
DR   Pfam; PF03797; Autotransporter; 1.
DR   Pfam; PF02395; IGA1; 1.
DR   Pfam; PF03212; Pertactin; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 1.
KW   Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT   SIGNAL         1      25       POTENTIAL.
FT   CHAIN         26    1008       IMMUNOGLOBULIN A1 PROTEASE.
FT   PROPEP      1009    1541       HELPER PEPTIDE (POTENTIAL).
FT   ACT_SITE     288     288       PROBABLE.
FT   MUTAGEN      288     288       S->T: LOSS OF ACTIVITY.
SQ   SEQUENCE   1541 AA;  169370 MW;  CB7257CB3196CC600 CRC64;

Query Match          2.4%;   Score 223;   DB 1;   Length 1541;
Best Local Similarity  19.3%;   Pred. No. 0.00031;
Matches   324;   Conservative  203;   Mismatches  589;   Indels  566;   Gaps   87;
```

```
Qy  271  DVHY----TDEAE--------VGTTNASIKAFIHKKYNDGLAASAAPVFLALVGDTDVISGEK  321
Db   31  DVDYQIFRDFAENKGKFSVGATNVLVK----DKNNKDL-GTALPNGIGPMI-DFSVVDVDK    84

Qy  322  GKK-TKKKVTDLYYSAVDGD---------YFPEMYTFRM  349
Db   85  RIATLINPQYVVGVKHVSNGVSELHFGNLNGNMNNGNAKAHRDVSSEENRYF----------  136

Qy  350  SASSPEELTNIIDKVLMYEKAT---------MP--DKSYLEKVLLIAGADYS----WNSQV   395
Db  137  SVEKNEYPTKLNGKTVTTEDQTQKRREDYYMPRLDKFVTEVAPIEASTASSDAGTYNDQN    196

Qy  396  GQPT---IKYGMQYYNQEHGYT---------DVYNY--LKAPYTGCYSHLN          433
Db  197  KYPAFVRLGSGSGSQFIYKKGDNYSLILNNHEVGGNNLKLVGDAYTYG-IAGTPYK--VNHEN 254

Qy  434  TG--VSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEV  492
Db  255  NGLIGFGN-----SKEEHSDPKGILSQ-DPLTN---YAVLG---------DSGSPLF---  293

Qy  493  ITRVKBKGAYAYIGSSP----NSYNGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFL    546
Db  294  -VYDREKGKWLFLGSYDFWAGYNKKSWQE---WNI----------YKSQFT           330

Qy  547  EDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNTYT  606
Db  331  KDVLNKDS----AGSL------IGSKT----DYSWSS-----NG----KTSTIT       361

Qy  607  LPASLPQNQASYSIQASAGSVVAISKDGVLYGTGVANASGVATVSMTKQITE--------N  659
Db  362  -----GGEKSLNVDLADG-----KDKPNHGKSV-TFEGSGTLTLNNNIDQGAGGLFFE    408

Qy  660  GNYDVVITRSNVLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPS---AKKAEG  715
Db  409  GDYEVKGTSDNTTWKGAGVSVAE-----------GKTVTWKVHNPQYDRLAKIGKG      453

Qy  716  SREVK------RIGDGLFVTIEPAN----------DVRANEAKVLAADNVWGDNTG     756
Db  454  TLIVEGTGDNKGSLKVGDGTVILKQQTNGSGGHAFASVGIVSGRSTILVLNDDKQVDPNSI 513

Qy  757  YQFLLDADHNTPGSVIPATGPLFTGTASSNLYSANFEYLVPANADP-------VVTTQN   808
Db  514  Y-----FG----FRGGRLDLNGNSLTFDHI--RNIDDGARLVNHNMTNASN           553

Qy  809  IIVTGQGEVVIPGGVYDYCITNPE---PASGKMWIAGDGGNQPARYDFTFEAGKKYTFT   865
Db  554  ITTTGESLITDPNTITPYNIDAPDEDNPYAFRR--IKDGGQLYLNLENYTYYALRK----  607

Qy  866  MRRAGMGDGTDMEVEDDSPASYYTVYRDGTKIK------------EGLTATTFEEDG     911
Db  608  -------GASTRSELPKNSGESNENWLYMGKTSDEAKRNVMNHINNERMNGFNGYFGEEEG  661

Qy  912  VAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNE-----FAPVQNLTGSVGQKVTLKWDA   966
Db  662  KNNGN------------------LNVITFKGKSEQNRFLLTGGTNLNGDLTVEKGTL--F  700

Qy  967  PNGTPNPNPNPNPNPGT------------KTIDADGDHGWKPGNA               1014
Db  701  LSGRPTPHARDIAGISSTKKDPHFAENNEVVVEDDWINRNFKATTMNVTGNASLYSGRN-  759

Qy 1015  PGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWCAQDANYASEHY  1074
Db  760  --VANITSN--ITASNKAQVHIGYKTGDT------------VCVR------SDY       791

Qy 1075  AVTASSTGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRH 1134
Db  792  TGVVTCT--------TDKLSDKALNS--FNPTNLRGNVNLT--ESANFVLG-KANLFGT  837

Qy 1135  FQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQ  1194
Db  838  IQSRGNSQVRLTE---------NSHWHLTGNSDVH------------QLDLANGH      871

Qy 1195  LDWLTAHGGSNV--------VSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVN-DGPPG 1246
Db  872  IHLNSADNSNNVTKYNTLTVNSLSGNG------------SFYYLTDLSNKQG         911

Qy 1247  DHYAVMISKTGTNAGDFTVVFERTPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAG  1306
Db  912  D--KVVVTKSAT--GNFTLQVADKTGEPNHNELTLFDASKAQRDHLNVSLVGNTVDLGA-  966

Qy 1307  TKY------------VAFRHYNCSDLNYILLDDIQFTMGGSP----------         1336
Db  967  WKYKLRNVNGRYDLYNPEVEKRNQTVDTTNITTPNNIQADVPSVPSNNEEIARVDEAPVP 1026

Qy 1337  ----TPTDYTYTVYRDG---------TKIKEGLTETTFEEDGVA---------TGNHEYC  1374
Db 1027  PPAPATPSETTETVAENSKQESKTVEKNEQDATETTAQNREVAKEAKSNVKANTQTNEVA  1086

Qy 1375  -------VEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAP--NSMDAILKWNA     1421
Db 1087  QGSGSETKETQTTETKETATVEKEEKAKVETEKTQ--EVPKVTSQVSPKQEQSTVQPQAE 1144

Qy 1422  PASKRAEVLNEDFENGIPASWKTIDADGDGNWTTTPPPGGSSFAGHNSAICVSSASHIN  1481
Db 1145  PARENDFTVN------------IKEPQSQTNTTADTEQPAKETSSNVEQPVTESTTVNTG 1192

Qy 1482  FEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASE---------HYAVYASSTGND  1531
Db 1193  NSVVENEN--TTPATTQP--------TVNSESSNKPKNRHRRSVRSVPHNVEPATTSSND 1243

Qy 1532  ASNFA---------NALLEEVLTAK--TVVTAPEAIRGTRAQ---------GTWYQKTV  1570
Db 1244  RSTVALCDLTSTNTNAVLSDA-RAKAQFVALNVGKAVSQHISQLEMNNEGQYNVWVSNT- 1301

Qy 1571  QLPAGTKYVAFRHFGC---------TDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETT 1626
Db 1302  SMNKNYSSSQYRRFSSKSTQTQLGW------DQTISNNVQLGGVFTYVRNSNNFDKATSKNT  1357

Qy 1627  YRDPDLATGFYT---------YGVKVVYPNGESAIET------ATLNITSLADVTAQKPYTLTVVG 1677
Db 1358  LAQVNFVSKYYADNHWYLGIDLGYGKFQSKLQTNHNAKFARHTAQFGLTAGKAFNLGNFG  1417

Qy 1678  KT 1679
Db 1418  IT 1419
```

```
RESULT 10
P2P_LACLC
ID   P2P_LACLC      STANDARD;      PRT;   1902 AA.
AC   P15293;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   PII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
DE   associated serine proteinase) (LP151).
GN   PRT.
OS   Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG   Plasmid pLP763.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX   NCBI_TaxID=1359;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCDO 763;
RX   MEDLINE=89313288; PubMed=2501630;
RA   Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirashima A.;
RT   "Molecular characterization of a cell wall-associated proteinase gene
RT   from Streptococcus lactis NCDO763.";
RL   Mol. Microbiol. 3:359-369(1989).
CC   -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC       GROWTH OF THE BACTERIA ON MILK.
CC   -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC       specificity, although some subsite preference have been noted,
CC       e.g. large hydrophobic residues in the P1 and P4 positions, and
CC       Pro in the P2 position. Best known for its action on caseins,
CC       although it has been shown to hydrolyze hemoglobin and oxidized
CC       insulin B-chain.
CC   -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC       an amide bond (Potential).
CC   -!- SIMILARITY: Belongs to peptidase family S8.
CC   -------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------
DR   EMBL; X14130; CAA32350.1; -.
DR   PIR; S06997, S06997.
DR   HSSP; P00782; 2SBT.
DR   MEROPS; S08.019; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
KW   Signal; Plasmid.
FT   SIGNAL        1     33       POTENTIAL.
FT   PROPEP       34    187       POTENTIAL.
FT   CHAIN       188   1870       PII-TYPE PROTEINASE.
FT   PROPEP     1871   1902       REMOVED BY SORTASE (POTENTIAL).
FT   ACT_SITE    217    217       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    281    281       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    620    620       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   SITE       1867   1871       LPXTG SORTING SIGNAL (POTENTIAL).
FT   MOD_RES    1870   1870       AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ   SEQUENCE 1902 AA; 200139 MW; 4B8D8B844D88CDF7 CRC64;
```

Query Match   2.4%;   Score 220.5;   DB 1;   Length 1902;

```
Best Local Similarity  19.5%;  Pred. No. 0.00058;
Matches  360;  Conservative  200;  Mismatches  672;  Indels  615;  Gaps   91;

Qy  160 MLGTMRGVRIAALTINPVQYDVVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFE 219
          :||: .||:  :::  .:   :.  ::  ::
Db   12 LAGTVALGALAVLPVGEIQAKAAISQQTKGSSLANTVTAATAKQAATDTTAAT------- 65

Qy  220 TAYKQLFNRDVYTDHGDLYNTPVRMLVVAGAKFKEALKPWLTWKAQKGFYLDV---HYT 275
          | ||   :.| :.| ||       ||        :| .| :|  :|:|
Db   66 ------NQAIAT-------QLAAKGIDYNK-----LNKVQQQDIYVDVIVQMSAA 102

Qy  276 DEAEVGT--TNASIKAFIHKKYNDGLAASAAPVFLALVGDTDVISGEK--------- 321
          .:|:    |.     |.:
Db  103 PASENGTLRTDYSSTAEIQQETNKVIRAQAS-VKAAVEQVTQQTAGESYGYVVNGFSTKV 161

Qy  322 ------------------------GKKTKKVTDLYY-- 341
                                  :|||| :  :||:::
Db  162 RVVDIPKLKQIAGVKTVTLAKVYYPTDAKANSMANVQAVWSNYKYKGEGTVVSVIDSGID 221

Qy  342 PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK 401
          ||   |   | :.|| :|| :     :|:|:
Db  222 PTHKDMRLSDDKDVKLT--------------KSDVEKFTDTAKHGRYFNSKV-----P 260

Qy  402 YGMQYYN---------QEHGYTDVINYLKAPYTGCYSHLNTGVSFANYTAH--GSETA 449
          || :||                : .||:|
Db  261 YGFNYADNNDTITDDTVDEQHG---------------MHV-AGIIGANGTGDDPAKSVV 303

Qy  450 WADPLLITSQLKALTNKDKYFLAIGNCCITAQF------------DYVQPCFGE------ 491
          .| |  |:  .:|  |:|:
Db  304 GVAPEAQLLAMKVFTNSDT-SATTGSATLVSAEDSAKIGADVLNMSLGSDSGNQTLEDP 362

Qy  492 ---VITRVKEKGAYAYI-----GSSPNSYWG--EDYY------------- 529
             ::  ::.:| ||        |   ::    : |:
Db  363 ELAAVQNANESGTAAVISAGNSGTSGGATEGVNKDYYGLQDNEMVGTPGTSRGATTVASA 422

Qy  530 QPT----------FEGT----------------SMGSYDATPLEDSYNTVNSIMWAGNLA 563
          ||:          |.|:                 .   :
Db  423 ENTDVITQAVTITDGTGLQLGPETIQLSSNDFTGSFDQKKPFVVKDA--SGNLSKGKVAD 480

Qy  564 --ATHAGNI----GNITHIGAHYYWEAYHVLG---------DGSVMPYRAMPKYNTYTLPAS 610
            ||.  :     ||||:|
Db  481 YTADAKGKIAIVKRGELTFADKQKTAQKAQAAGAGLIIVNNDGTATPVTSMALTTTF--- 535

Qy  611 LPQNQASYSIQASAG----SYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVI 666
          .:       :  |:      :|:
Db  536 ------------PTFGLSSVTGQKLVDWVAAHPDDSL-GVKIA----LTLVPNQKYTED-- 577

Qy  667 TRSNYLPVIKQIQVGEPSPYQPVSNLT----ATTQGQKVTLKWEAPSAKKAEGSREVKRI 722
          |    :    :  : :|   |        |
Db  578 ---------KMSDFTSYGPVSNLSFKPDITAPGGNI---WSTQN------------ 609

Qy  723 GDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGT 782
          ||.|  :|       .   | :  :
Db  610 ------------------NNGYTNMSGTSMASPFI---------------- 654

Qy  783 ASSNLYSANFEYLVPAN-ADPV--VTTQNIIVT---GQGEVVIPGGVYDYCITNPEPAS 835
          |:   : |  |:|       .:    ::
Db  655 ALT-----DFLKTVEMNTAQPINDINYNNVVSPRRQGAGLVDVKAAI-DALEKNPST-- 706

Qy  836 GKMWIAGDGGNQPARYDDFT-FEAGKKYTFTMRRA----GMGDGTDMEV----EDDSPA 885
          || .| |||| :    |       .|| |
Db  707 ----VVAENGYPAVELKDFTSTDKTFKLIFTNRTTHELTYQMDSNTDTNAVYTSATDPNS 762

Qy  886 SYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVFGSNEF 945
          ::   :: : .|  :   :   .      :::  :
Db  763 GVLYDKKIDGAAIKAGSNIT----VPAGK---TAQIEFTLSLPKSFDQQQFVBG----- 809

Qy  946 APVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASW---KTIDA 1002
          |           |:                         :
Db  810 --FLNFKGSD-GSRLNLPY----------------MGFPGDWNDGKIVD-- 839

Qy 1003 DGDGHGWKPGNAPGIAGYNSNGCVYSESPGLGGIGVLTPDNYLITPALDLPNGGKLTFWV 1062
          :    ||:    ::.:.:
```

```
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=1577609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
RN   [3]
RP   SEQUENCE OF 1-765 FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=91100301; PubMed=1987123;
RA   Wu J.Y., Siegel L.M., Kredich N.M.;
RT   "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT   requirement for a cloned cysG plasmid to overcome limiting siroheme
RT   cofactor.";
RL   J. Bacteriol. 173:325-333(1991).
CC   -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC       in positions 414 and 732.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF133696; AAD39458.1; -.
DR   EMBL; AE008859; AAL22340.1; -.
DR   EMBL; M64606; AAA27042.1; ALT_FRAME.
DR   EMBL; M64606; AAA27043.1; ALT_FRAME.
DR   PIR; C39200; C39200.
DR   PIR; D39200; D39200.
KW   StyGene; SG10437; bigA.
KW   Virulence; Repeat; Signal; Complete proteome.
FT   SIGNAL         1      27       POTENTIAL.
FT   CHAIN         28    1953       PUTATIVE SURFACE-EXPOSED VIRULENCE
FT                                  PROTEIN BIGA.
FT   DOMAIN       101     252       15 X 11 AA TANDEM REPEATS.
FT   REPEAT       101     103       1 (INCOMPLETE).
FT   REPEAT       104     113       2 (INCOMPLETE).
FT   REPEAT       114     122       3 (INCOMPLETE).
FT   REPEAT       123     133       4.
FT   REPEAT       134     144       5.
FT   REPEAT       145     155       6.
FT   REPEAT       156     166       7.
FT   REPEAT       167     177       8.
FT   REPEAT       178     188       9.
FT   REPEAT       189     199       10.
FT   REPEAT       200     210       11.
FT   REPEAT       211     221       12.
FT   REPEAT       222     232       13.
FT   REPEAT       233     243       14.
FT   REPEAT       244     252       15 (INCOMPLETE).
FT   CONFLICT     207     207       D -> DRGDDDVTPPDD (IN REF. 1).
FT   CONFLICT     514     514       A -> R (IN REF. 3).
FT   CONFLICT     698    1698       D -> N (IN REF. 1).
FT   CONFLICT    1795    1798       QYLE -> ITLQ (IN REF. 1).
FT   CONFLICT    1836    1837       SA -> T (IN REF. 1).
SQ   SEQUENCE  1953 AA;  200150 MW;  61B3F1C954D91AE CRC64;

Query Match             2.4%;
Best Local Similarity   20.2%;   Score 219.5;  DB 1;  Length 1953;
Matches  266;  Conservative 125;  Mismatches 408;  Indels 521;  Gaps 70;

Qy      663  DVVITRSNYLPVIK---------------------QIQVGEPSPYQPVSNLTATTQGQKVTLK--- 704
              |||  ||  |  |                         ||| |
```

```
Db    840  ----------------------------SLNGITYSPAGG----------NFGTVPLLTNKNTGTQYYGG  871

Qy   1063  CAQDAN------------YASEHYAVYASSTGNDAS------------NFTNALLEETITAKGVRS  1104
                         :|:||:|||||                        :|:|
Db    872  MVTDADGNQTVDDQAIAFSSDKNALY------NDISMKYYLLRNISNVQV-DILDGQQ--  922

Qy   1105  PKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFE  1164
                  ||| |  | | |:||  :|:|   :| |:||
Db    923  -NKVTTLSSSTNRKKTYYNAHSQQYI-YHAPAWDGTYYDQRDGNIKT--ADDGSYTYR  977

Qy   1165  SSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLI  1224
                           |:|| |        :|:|
Db    978  ISGVPEGGDKRQVFDV-----------PFKLLDSKAPTVRHVALSAKTENG--KTQYYIT  1023

Qy   1225  SK---DVTG--ATK-VKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP--------  1271
             ::    :|:||||||||||||||||||
Db   1024  AEAKDDLSGLDATKSVK--TAINEVTNLD--ATFTDAGTTADGYTKI--ETPLSDEQAQ  1076

Qy   1272  ---NGINKGGARFGLSTEANGAXPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDI  1328
                  |||  | |:||      ::
Db   1077  ALGNGDN-SAELYLTDNASNATDQ---DASVQKPGSTSF------------------DL  1113

Qy   1329  QFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKC  1388
                                          ::::
Db   1114  IVNGGGIPDKISSTTTGYEANT---QGGGTYTF-----SGTYPAAVDGTYTDAQGKKHD  1164

Qy   1389  VDVTVNS------TQFNPVQN-------LTAEQAPNSM---DAILKWNAPASKRAEVLN  1431
                  :: :                                        |
Db   1165  LNTTYDAATNSFTASMPVTNADYAAQVDLYADKAHTQLLKHFDTKVRLTAPT-----FTD  1219

Qy   1432  EDFENGI-PASWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDN  1490
                                                      : :
Db   1220  LKFNNGSDQTSEATIKVTGTVSADTKTVNVG------DTVAALDAQHHFSVDVPVN---  1269

Qy   1491  YLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYAS--------STGNDASN  1534
                                                              :
Db   1270  ----------YGDNTIKVTATDEDGNTTTEQKITTSSYDPDMLKNSVTFDQGVTFGANEFN  1320

Qy   1535  FANALLEVLITAKTVVTAPEAIRGTRAQ------------------GTWYQKTVQLPA  1574
                                                              ||||||||
Db   1321  ATSAKFYDPKTGIATITGKVKHPTTTLQVDGKQIPIKDLFTFSFTLDLGTLGQKPFGVVV  1380

Qy   1575  G--TKYVAFRHFGCTDFFWINLDDVVIT------SGNAPSYTYTIYRNNTQIASGVTET--  1625
                                                              :|
Db   1381  GDTTQNKTFQ-----EALTFIILDAVAPTLSLDSSTDAPVYT--NNPNFQITGTATDNAQ  1432

Qy   1626  ----TYRDPDLATGFYTYGVKVVYPNGESAIE-------TATLNITSLADVTAQKPY  1671
                  |                                            |
Db   1433  YLSLSINGSSVASQYVDININSGKP-GHMAIDQPVKLLEGKNVLTVAVTDSEDNTTTKNI  1491

Qy   1672  TL------TVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVTAQGG  1712
                                      :
Db   1492  TVYYEPKKTLAAPTVIPSTTEPAKTVLTLTANSAATG-ETVQYSADGG  1537
```

```
RESULT 11
BIGA_SALTY
ID   BIGA_SALTY     STANDARD;      PRT;   1953 AA.
AC   P25927; P25928; Q9XCQ3;
DT   01-MAY-1992 (Rel. 22, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative surface-exposed virulence protein bigA precursor.
GN   BIGA OR STM3478.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 14028;
RA   Stojiljkovic I., Valentine P., Heffron F.;
RT   "Salmonella typhimurium rhs homolog.";
```

```
Qy   1548 TVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTD---FFWINL------ 1594
Db   1189 -------TVNG--------TVTINGGAN-ALANYGTLDADAISTWHSLFNEADGSIT 1229

Qy   1595 DDVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPPLATG------FYTY 1639
Db   1230 TDLLTLNGDVTFY------NNGDFTGSIAGTSYQQEIVNTGDMTVAEDGKSLVSGSFYFY 1283

Qy   1640 GVK-VVYPNGESAI--ETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNG 1695
Db   1284 NEEDATLTNSGSAVEGGENTINLTRANDSLTQ------VNSGTITAT-NGYSAITTVNG 1336


RESULT 12
YS89_CAEEL
ID   YS89_CAEEL     STANDARD;      PRT;   3178 AA.
AC   Q09624; Q09625; Q969D4;
DT   01-NOV-1995 (Rel. 32, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein ZK945.9 in chromosome II.
GN   ZK945.9/ZK945.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Wilkinson-Sproat J.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Durbin R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   -!- SIMILARITY: Contains 1 PLAT domain.
CC   --------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------
DR   EMBL; Z48544; CAB70192.1; -.
DR   EMBL; Z48582; CAB70192.1; JOINED.
DR   EMBL; Z48582; CAB70201.1; -.
DR   EMBL; Z48544; CAB70201.1; JOINED.
DR   WormPep; ZK945.9; CE25697.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR005821; Ion_trans.
DR   InterPro; IPR001024; Lipoxygenase_LH2.
DR   InterPro; IPR003915; PKD_2.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   InterPro; IPR008976; PLAT_LH2.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   Pfam; PF01477; PLAT; 1.
DR   PRINTS; PR01433; POLYCYSTIN2.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00308; LH2; 1.
DR   PROSITE; PS50095; PLAT; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM       13     30       POTENTIAL.
FT   TRANSMEM       51     73       POTENTIAL.
FT   TRANSMEM     2139   2161       POTENTIAL.
FT   TRANSMEM     2348   2367       POTENTIAL.
FT   TRANSMEM     2390   2412       POTENTIAL.
FT   TRANSMEM     2451   2468       POTENTIAL.
```

```
Db    251 DSVITFSNGVTIDKGKDTLITFDSFKLDNGSVLEGAVWNYSEQDNQWQLITADGKTLNVTG 310

Qy    705 WEAPSAKKA--EGSREVKRIGDGLF--------------VTIEPANDVRANEAK------- 742
Db    311 WDVTDANAAVIEGTQE-----NGLYWKYDSRGYLIIADDNTTVISGDDQAHNSDRGMDIS 365

Qy    743 ------VVLAADNVWGDNTGYFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEY- 794
Db    366 --------ATGMVISGDGTTNTISGHSTVDN 415

Qy    795 --LVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPAR 850
Db    416 ATGALISGNG----TTTNF---AGDIAVSGGGTAIII--------DGDN----- 449

Qy    851 YDDFTFEAGKKYTFMRRAGMG--------DGTDMEVEDDSPASYTYTVYRDGTKIKEGLTA 904
Db    450 ---ATIKNTGTSDISGAGSTGTVIDGNNARVNNDGDMTIT--------DGGTG 491

Qy    905 TTFBEDGVAAGNHEYCVEVKYTAGVSPKVCKDVT--VEGSNEFA---PVQNLTGSSVGQ 958
Db    492 GHITGDNVVIDN-----AGSTTVSGADATALYIEGDNALVINEGNQTISGGAVGT 541

Qy    959 KVTLKWDAPNGTPNENPNENPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIA 1018
Db    542 RI----------------DGDD----------A 548

Qy   1019 GYNSNGCVYSESFGLGGIGVLITPDNYLITPALDL--PNG--GKLTFWVCAQDANYASEHY 1074
Db    549 HTTNTGDIAVDGAGSAAV-IINGDNGSLTQAGDLLVTDGAMGIITY--------- 593

Qy   1075 AVYASSTGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQK-TVDLPA-GTKYVAF 1132
Db    594 ----GTGNEAKNTGNA------TVRDADSVGFVVAGEKNTFKNKGDIDVSLNGTGALV- 641

Qy   1133 RHFQSTDMFYIDLD-EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGD--------G 1184
Db    642 ----SGDMSQVTLDGDINVVS---VQDSEGVFSSATGVSVSGDSNAVDITGNVNISADYG 694

Qy   1185 QGWLCLSSGGLDWLTAHGGSNVVSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGF 1244
Db    695 QDDLAAGAPPLTGVVVGGNGNTV---TLNG-ALNIDDNDLS-----ATGGQYLDVVGLSV 745

Qy   1245 PGDHYAVMLSKTGTNAGDFTVVFEETP--------NGINKGGARFGLSTEANGAKPQSVW 1296
Db    746 TGDDNDVEID-----GGINITHSEDPLDGTSADITGISVSG--NSTVTLNG------- 789

Qy   1297 IERTVDLPAGTKYVAFRHYNCSDLN----YILLDDIQFTMGGSPTTPTD-YTYT--VYRDG 1349
Db    790 -HSTID---TNTVVGGHVVLARVNNGGSLILGDDSVVDVNVSYIPTGYYTYNALLMADG 844

Qy   1350 TKIKBGLITETTFBEDGVATGNHEYCVEVKYTAGVSPKKCVDTVNSTQFNPVQN------ 1403
Db    845 --EG---TSIENKGDITSHGVISV-IRADNGSEVSNSGDILVYATSSNSSEDRAAITR 896

Qy   1404 --------LTABQAPNSMDAI----LKW---------- 1419
Db    897 ASGEGSAVHNKAGGDITLISDQTPQGSGG-EVYPLKWYTHTFYAMMASDYGDVVNDEGAT 956

Qy   1420 ------NAPASKRAEVLNE----------DFENGIPAS--WK------------ 1443
Db    957 IHLQGAGVYGVTASRGKALNEGNIYLDGLVPTLDDENNITSTSYWQPSSLYLTSSGMVAG 1016

Qy   1444 TIDADGDG---------NN-----------WTTTPPG--------- 1461
Db   1017 STDADGDATAINTGNITVNNAGFGMMALNGGTAINQGVITLITADDGVTGQADELVGMAAL 1076

Qy   1462 --------GSSFAGHNSAICVSSASHINFEGP--QNPDNYL-VTPE--- 1496
Db   1077 NGGVVINDTSGVINIDADYGQAFLSDSSSYIINNGS-INLNGSPMDDTDSHMGGTPTDKI 1135

Qy   1497 --LSLPGGG---TLT----FWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAK 1547
Db   1136 WIQSLPGSGSDTRTSDTGFFTAGTLANYGTETLNGDVDVNGGWLYNEAGASL------- 1188
```

```
FT   TRANSMEM      2483   2505       POTENTIAL.
FT   TRANSMEM      2567   2589       POTENTIAL.
FT   TRANSMEM      2836   2858       POTENTIAL.
FT   TRANSMEM      2939   2961       POTENTIAL.
FT   TRANSMEM      2976   2998       POTENTIAL.
FT   TRANSMEM      3038   3060       POTENTIAL.
FT   DOMAIN         266   1196       SER/THR-RICH.
FT   DOMAIN        1105   1241       GLY/SER-RICH.
FT   DOMAIN        2071   2120       GPS.
FT   DOMAIN        2182   2305       PLAT.
SQ   SEQUENCE      3178 AA;  344726 MW;  F8239436D03666CD CRC64;

Query Match             2.4%;  Score 218.5;  DB 1;  Length 3178;
Best Local Similarity  19.8%;  Pred. No. 0.0016;
Matches  285;  Conservative 185;  Mismatches 525;  Indels 443;  Gaps  72;

Qy    426  TGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQPFDYV  485
               :: :||:|:  :     :| :    ::|::|: | :|:::::||  ||:
Db    501  TGPSSSSSTPSSTASSSV--SSTASSTQSSTSTQQSSTTTKSE---------TTTSSD--  547

Qy    486  QPCFGEVITRVKEKGAYAYIGSSPNSTWGE------DYYWSVGAN------AVFGVQPTFE  534
                ::  |: :   |       ::|      | : :    : ||    :
Db    548  -----GTNPDFYFVEKATTTFYDSTSVNLTLNSGLGIIGYQTSIE--------------  587

Qy    535  GTS--MGSYDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIG-AHYYWEAYHVLGDG  591
               ||      : : |:        ::   : :  |     : ::: :
Db    588  CTSPTSSNYVSTTKDGCACFTKSVSM--PRLGGTYPAS---TFVGPGNYTFRATMTTDDK  641

Qy    592  SVMPYRAMPKINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVS  651
            : :     :|:          :  :    :     :   |    :  |||:
Db    642  KVY------------YTYANVVIQEYSSTTIESBS------STSAVASSTSSTPSTPSSTLS  685

Qy    652  MTKQITENGNYDVVITRSNYLPVIKQIVGEPSPYQPVSNLTATTGQKVTLKWEAPSAK  711
               :::   | :  ::::|   :  ::|        ::: : :|  | |: :
Db    686  -TSTVIEPSS------TRS--------------SDSTTTSAGSTTTLQESTTTSE  719

Qy    712  KAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDAD--------  764
               ::| || |  :|       :|:|   :     :   ||:|:
Db    720  ESTTDSSTTTISDTSTTSSSPSSTTA--DSTSTLSVDQ-------FDFILDSGLSWNETR  770

Qy    765  HNTFG--SVIPATGPLFTGTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGV  823
            |||       :  |            |     ::  :| :  :  :    |
Db    771  HNEDSINIVPLPTNAITPTERSQTFECR----------NVSTEPFLIIKESTCL----N  815

Qy    824  YDYCIINPEPASGKMWIAGDGGNQPARYDDFTFEAG-KKYTFTMRRAGMGDGTDMEVEDD  882
             :  | :        : : :         | :    |    :    ::   ::
Db    816  YSNTVLNATYSS-----------NIPIQPIE-TFLVGIGTYEF--RINMTDLTTMQVVSH  861

Qy    883  SPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGS  942
               ::|: : :|   | |:       |   | | :         :   |     : :
Db    862  ---IFTLNVVADSTSTSE-VTSTT------STGSSSEBSSAISTTSGIE----STSTLEAS  907

Qy    943  NEFAPVQNLTGSS----VGQKVTLKWDAPNGTPNFNPNPNPNPGTTLSESFENGIPASWK  998
               ::    ::| :    ::  ::|      :  | : | :      : : :|  ||
Db    908  TTDASQDSSTTSTSDSGTTSDSTTI---PSTSDSSGLSQTPSDSSASDSMR  961

Qy    999  TIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNVLITPALDLPN--GG  1056
            :  || |      :      ::       | :|   :  | :   |  |     |
Db    962  TTTVDPDASTETPYDFV-LENLTWNETVY------------YSENPFYITP---IPNKEPG  1006

Qy   1057  KLTFWVCAQDANYASEHY------------------AVYASSTGNDASNFTNA------LLE  1094
            |:  |:      | :                    ||  :  |    :||
Db   1007  ALTTAMTCQCRNDSSQPFVLLKESNCLTEFGKNGAVSASVSFNPMTSFVPATGTYEFLIN  1066

Qy   1095  ETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHQSTDMFYIDLDEVEIKANG  1154
               : |:  |     | :| ||   | ||:| |:|||||||| ::: : : :  |
Db   1067  VTNRASGESASHIF------TMNVVLPTTTTETPTTVSSSD--------DAGG  1106

Qy   1155  KRADFTETFESSTHGEAPAEWTTIDADGD-----------GQGWICLSSGQLDWLTAHGGSNVV  1207
                    |:::|         ::                | |    :|     |   |:
Db   1107  KTGGTGAT--GGTGGTGSGGSGSATTLSTGDAVRSTTSGSG-----SGQSSTSGSGAGSGTT  1159

Qy   1208  SSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVF  1267
```

```
Db   1160  ASSGSSG---------------GSSGTGSD-----GVNSG-----KTTALNGDGTGSGTAT--  1195
                :||:|                : ||: |     ||:||     |:|  :|  |:  ||::
Qy   1268  EETPNG-INKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKVAFRHYNCSDLNYILLD  1326
               :||||                            :| ::|||::|:|:|::  ||:|||
Db   1196  --TPGSHLGDGGSTSGSGSGSDSNGSSGVS----------------------TK------  1223

Qy   1327  DIQFTMGGSPTPTDYTYTVYRDGTKIKEGL-------------TETTFEEDGVATGNHEYCV  1375
                 |: : |:                                 :: |:
Db   1224  ----SSSGSDT-------SGSSDSSGANGAFSATAQPSTRTTKTRSSLAT-----  1262

Qy   1376  EVKYTAGVSPKKCVDVTVNSTQFNPVQNLTA------------EQAPNSMDAILKW--NAPASK  1425
                ::   :|| |  : |: :|: |  :|:             : |   : :||:
Db   1263  ------VSPISAAEQAIIDAQKADVMNQLAGIMDGSASNNSLNTSSSLLNQISSLPAAD  1315

Qy   1426  RAEVLNEDFENGIPASWKTIDADGDGN------------NWTTTPPGGGSFAGHNSAIC  1473
               :  :|  :||:|||: ||:|| |  :              :::
Db   1316  LVEVAQSLLSN-------TLKIPGVGNMSSVDVLKTLQDNIATT------NSELADEMAKV  1363

Qy   1474  VSSASHINFEGPQNPDNVLIVTPELSLPGGGTLTFWVC----AQDANYA------SEHY  1521
                ::: : ::    :: : :  :| | : ||: :|:        ::  :      ::|
Db   1364  ITKLANVNMTSAQSLNSYLSSLDLALKGSTVYTLGVSSTKSKDGTYAVIFGYVIASGYTL  1423

Qy   1522  ------AVYASS--TGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAAQGTWYQKTVQ  1571
                     ||||:    |     :                    : |  |:  |::::|
Db   1424  VSPRCTLSIYGSTIYLTGDTRASY-KQLDGDTVTADTMLAAAIGIQGMFATN---GRTVQ  1479

Qy   1572  ------------------------LPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTYRNNTQIAS  1620
                                     :::   : : ::    : :  ::  ::  : ::
Db   1480  VEQDKIDDKRSLVSGNIMATMSGVGDVQSGEYSYNDMYVTAMNV---TY--DNSTVGS  1532

Qy   1621  GVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATL-----------NITSLADVTAQKPYTL  1673
               : :: :   |::   ||  :|  |:            :  ||| ||    |:
Db   1533  TSQKNT------------SFSFNIPVSEVQYILLIESGTMKLHSTQNIVSRGLVV---TA  1578

Qy   1674  TVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGK--SYVEKL  1729
               :||| ||||::     :    |                          :||||
Db   1579  SYGGVYTTITCT------NGTGKFVEVDTDNAIFSYNADSFTV-VASDGSSASTVKKL  1629
```

```
RESULT 13
APU_THETU
ID   APU_THETU               STANDARD;      PRT;   1861 AA.
AC   P38536;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Amylopullulanase precursor (Alpha-amylase/pullulanase) (Pullulanase
DE   type II) [Includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-glucan
DE   glucanohydrolase); Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan
DE   glucanohydrolase) (Alpha-dextrin endo-1,6-alpha-glucosidase)].
GN   AMYB.
OS   Thermoanaerobacter thermosulfurogenes (Clostridium
OS   thermosulfurogenes).
OC   Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC   Thermoanaerobacteriaceae; Thermoanaerobacterium.
OX   NCBI_TaxID=33950;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 3896 / EM1;
RX   MEDLINE=9252998; PubMed=8195085;
RA   Matuschek M., Burchhardt G., Sahm K., Bahl H.;
RT   "Pullulanase of Thermoanaerobacterium thermosulfurigenes EM1
RT   (Clostridium thermosulfurogenes): molecular analysis of the gene,
RT   composite structure of the enzyme, and a common model for its
RT   attachment to the cell surface.";
RL   J. Bacteriol. 176:3295-3302(1994).
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC       linkages in oligosaccharides and polysaccharides.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
CC       linkages in pullulan and in amylopectin and glycogen, and the
CC       alpha- and beta-limit dextrins of amylopectin and glycogen.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
```

```
CC   -!- SUBCELLULAR LOCATION: CELL-BOUND. IT C-TERMINUS MAY SERVE AS AN
CC       S-LAYER ANCHOR.
CC   -!- PTM: Glycosylated.
CC   -!- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
CC   -!- SIMILARITY: Contains 2 fibronectin type III domains.
CC   -!- SIMILARITY: Contains 3 s-layer homology (SLH) domains.
CC   ---------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
DR   EMBL; M57692; AAB00841.1; -.
DR   HSSP; Q08751; 1BVZ.
DR   InterPro; IPR006589; Alp_amyl_cat_sub.
DR   InterPro; IPR006048; Alpha_amyl_C.
DR   InterPro; IPR006047; Alpha_amyl_cat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR004185; Glyco_hydro_13Ig.
DR   InterPro; IPR004193; Glyco_hydro_13N.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00128; alpha-amylase; 1.
DR   Pfam; PF02806; alpha-amylase_C; 1.
DR   Pfam; PF02903; alpha-amylase_N; 1.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF02922; isoamylase_N; 1.
DR   Pfam; PF00395; SLH; 3.
DR   SMART; SM00642; Aamy; 1.
DR   SMART; SM00632; Aamy_C; 1.
DR   SMART; SM00060; FN3; 2.
DR   PROSITE; PS01072; SLH_DOMAIN; 3.
KW   Carbohydrate metabolism; Multifunctional enzyme; Hydrolase;
KW   Glycosidase; Calcium-binding; Repeat; Signal; Glycoprotein.
FT   SIGNAL          1     35       POTENTIAL.
FT   CHAIN          36   1861       AMYLOPULLULANASE.
FT   DOMAIN        928   1018       FIBRONECTIN TYPE-III 1.
FT   DOMAIN       1157   1248       FIBRONECTIN TYPE-III 2.
FT   ACT_SITE      628    628       BY SIMILARITY.
FT   ACT_SITE      657    657       BY SIMILARITY.
FT   ACT_SITE      734    734       BY SIMILARITY.
FT   METAL         401    401       CALCIUM (BY SIMILARITY).
FT   METAL         403    403       CALCIUM (VIA CARBONYL OXYGEN) (BY
FT                                  SIMILARITY).
FT   METAL         406    406       CALCIUM (BY SIMILARITY).
FT   METAL         407    407       CALCIUM (BY SIMILARITY).
FT   METAL         452    452       CALCIUM (VIA CARBONYL OXYGEN) (BY
FT                                  SIMILARITY).
FT   METAL         453    453       CALCIUM (BY SIMILARITY).
FT   DOMAIN       1681   1739       SLH 1.
FT   DOMAIN       1740   1803       SLH 2.
FT   DOMAIN       1804   1861       SLH 3.
FT   CONFLICT     1734   1734       D -> E (IN REF. 1; AAB00841).
SQ   SEQUENCE   1861 AA;  206104 MW;  06C23070E5453B574 CRC64;
```

```
Query Match             2.3%;  Score 214;  DB 1;  Length 1861;
Best Local Similarity  18.8%;  Pred. No. 0.0013;
Matches  332;  Conservative 200;  Mismatches 638;  Indels 598;  Gaps 87;

Qy    139 VVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYDVVANQLKVRNNIEIEVSF 198
              ||:|  |:| |:|| | ||:| | |  |: :|  :|  |||: |||:||:::::
Db    178 VYSYTAHVPKGDYQYKVT----LG----------NTWDENYGANGVKDGSNIQINVT-- 220

Qy    199 QGADEVATQRLYDA------SPSPYFETAYKQLFNRDVYTD-HGDLYNTPVRMLVVAGA 250
              :: |           | |    ||:|: || :| | ::  ||:  |   ::
Db    221 --NDADITFYDANTHNIWTNVSPILTGLDNNIYYDDLKHDTHDSFFRNPFGAVKV-- 274

Qy    251 KFPKEALKPWLTWKAQ-KGFYLD---VHYTDE----AEVGTTNA---------SIK-A 289
                ::  | : |    |  :||:   |: |     |:  :||  :
Db    275 ----DQTVTLRIQAKNHDLESARISYWDDINKIRIELPMTRIGESPDGNFEYWEIKLS 328
```

```
Qy    290 FIHKK------YNDGLAASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFP 342
              ||:                                 ||  |
Db    329 FDHPTRIWYFILKDGYKT----------AYYGDNDDQLGGVGKAT--------DTVNKDF- 371

Qy    343 EMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKY 402
              |: |: :      :| ||:        | ::| || | :|:: |
Db    372 ELTVYDKNFDTPDWMKGAV----MYQ-IFPDRFYNGD------TSNDHAKTLSRGNDPIE- 420

Qy    403 GMQYYYNQEHGYTDVINYLKAP-YTGCYSHLNTGVSFANYIAHGSETAWADPLLTTSQLK 461
              |  :   |||:|||:|| |  |||: ||:|||||:: :|| | :|  |
Db    421 ----FHNNWNDLPDNPNNAGTPGYTG------------------DGIWSNDFF--GDLK 455

Qy    462 ALTNKDKYFLAIGNCCI------------TAQFDYVQPCFG------EVITRVKEKG 500
              ::: :  |||: ||                   ::    :::::
Db    456 GIDDKLDYLKGLGVSVIYLNPIFESPSNHKYDTADYTKIDEMFGTTQDFEKLMSDAHAKG 515

Qy    501 AYAYIGSSPNSYWGEDYY------------WSVGANAVFGVQPTFEGTSMGSYDATF 545
              :  |  :  :  ||:::                 ||:  | |:| ||:  :|:
Db    516 IKIILDGVFNHTSDDSIYFNRYGKYPGLGAYQAWKEGNQSLSPYGDWYTINSDGTYECWW 575

Qy    546 LEDSYNTVNSI------------MWA-----------GNLAATHAGNIGNITHIGAH 579
              ||: |  |                                :|:|  ||:||: |:
Db    576 GYDSLPVIKSLNGSEYNVTSWANFIINDENAISKYWLNPDGNLNDGADGWRLDVENEVAH 635

Qy    580 YYWEAYH------------------------VLGD-GSVMPY 596
              ||:                               ||| |:|
Db    636 DFWTHFRNAINTVKFEAPMIAENWGDASLDLLGDSFNSVMNYQFRNDIIDFLIGQSPDDG 695

Qy    597 -------------RAMPKINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVA 642
                                 |||:||||: |  ||: || :| ||| |
Db    696 NGQHNPIDAAKLDQRLMSIYERYPLP-----AFYSTMNLLGSHDTMR---ILTVFGYN 745

Qy    643 NASGVATVSMTKQITENGNYDVVITRSNYLPVIKQIQVGEPSPY---QPVSNLIATTQG 698
              |:   |: |  |:  ||:  :|:|||||:| ||:|  |||:|      |: :|  |
Db    746 SADFNENSDAAKQLAEQKLKLATILQMGY-PGMADIYYGDEAGVSGGKDPDDRRTFPWGN 804

Qy    699 QKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQ 758
              ||:| || | ||  ||::: |:| :| :|  || | |  || :  | |::| |:: :|
Db    805 EDTTLQDFFKNISSIRNNNQVLKTGD--LETLYAQNDVYAIGRRIINGKD--AFGTSYP 859

Qy    759 ------FLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLVPANADPVVTTQNIIVTG 813
                    ||  |:|| |:|  ||  ||| :::||: ||:||:|| | |||| ||:|::|
Db    860 DSAAIVAINRSKSDKQIAIDTTKFLRDGVTFKDLINNNVSYSI-----------------S 903

Qy    814 QGEVVIPGGVYDYCITNPEPA-SGKWWIAGDGNQOPARYDDFTFEAGKKYTFTMRRAGMG 872
              |:| ||  | |||: |:    |||:  :||| :||  :    |:    | |:|   |:
Db    904 NGQIVI-----------------DVPAMSGVMLISDDG------QDLTAPQAPSNVVVTSGNGKV 945

Qy    873 DGTDMEVEDDSPASYTYTVRDGT----------KIKEGLTATTFEEDGVAAG-NHEYCEVKY 925
              |  |::  |: :|:|||||:|              :|| ||:||:|:|  |:|||| |::
Db    946 DLSWLQ---------SDGATGYNIYRSSVEGGLYEKIASNVTETTFEDANVTNGLKYVVAISAID 1001

Qy    926 TAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAP----------NGTPNPNP 975
              || :||  |::|:| :|:  | :| ||  |:||:|  ||
Db   1002 ELGNESGISNDAVAYPAYPIGWVGNLT--QVSDNHIIGVDKPTEDIYAEVWADGLTN-ST 1058

Qy    976 NPNPNPGTTLSESFENG------IPASWKTI-DADGDHGWKPGNAPGIAGYNSNGCVYS 1028
              ||||| |||:|| :||               || || | || ||  |  |  :| :||
Db   1059 GQGPNMIAQLGYKYVSGTVYDSVYGSVVNSVYGVDDSGFTWVNAQYVGDIGNNDQ---YK 1115

Qy   1029 ESFGLGGIGVLTPD--------------NYLI--TPALDLPNGGKLTFWVCAQD----ANYASE- 1072
              || :|  || : |               |:|   || | ||| ::|:|: || |      |||:
Db   1116 ASF---------------TPDKIGQWEYLMRFSDNQGQDWITTSTLSFYVVPSDDLIKPTAPYLNQP 1167

Qy   1073 -----------------HYAVYASSTG--NDASNFTNALLE--ETITAKGVRSPKA 1107
                                :| ::  ||  |::|:|   ||:  :| ||  |: |:
Db   1168 GTESSRVSLTWNPSTDNVGIYDYEIYRSDGGTFNKIATVSNEVVNYIDTSVINGV---- 1222

Qy   1108 IRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLLDEVEIKA--NGKRADFTE-- 1161
              : ||:  || | ::|: |  :|:|| |: |:  ||: | || |  |    :|
Db   1223 --TYNYKVVAVD----LSFNRTES-NVVTIKPDVVPIKVIFNVTVPDYTPDAVN 1269

Qy   1162 --TFESSTHGEAPAEWTIDADGDGQWLCLSSG--QLDWLTAHGG- 1203
```

```
CC   -!- SIMILARITY: Belongs to the intimin/invasin family.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AE000289; AAC75042.1; ALT_INIT.
DR   EMBL; D90837; BAA15800.1; -.
DR   EMBL; D90836; BAA15799.1; ALT_INIT.
DR   EcoGene; EG13378; yeeJ.
DR   InterPro; IPR003344; Big_1.
DR   InterPro; IPR003535; Intimin.
DR   InterPro; IPR008964; Invasin_intimin.
DR   InterPro; IPR002482; LysM.
DR   InterPro; IPR000601; PKD.
DR   Pfam; PF02369; Big_1; 13.
DR   PRINTS; PR01369; INTIMIN.
DR   SMART; SM00634; BID_1; 13.
DR   SMART; SM00257; LysM; 1.
DR   SMART; SM00089; PKD; 6.
KW   Hypothetical protein; Repeat; Complete proteome.
FT   DOMAIN      738    834       BIG-1 1.
FT   DOMAIN      840    931       BIG-1 2.
FT   DOMAIN      932   1033       BIG-1 3.
FT   DOMAIN     1042   1137       BIG-1 4.
FT   DOMAIN     1146   1237       BIG-1 5.
FT   DOMAIN     1246   1350       BIG-1 6.
FT   DOMAIN     1351   1448       BIG-1 7.
FT   DOMAIN     1449   1553       BIG-1 8.
FT   DOMAIN     1554   1655       BIG-1 9.
FT   DOMAIN     1661   1754       BIG-1 10.
FT   DOMAIN     1763   1853       BIG-1 11.
FT   DOMAIN     1855   1950       BIG-1 12.
FT   DOMAIN     1952   2053       BIG-1 13.
FT   CONFLICT    105    105       S -> G (IN REF. 2).
SQ   SEQUENCE   2358 AA;  248599 MW;  232249750BF631ED CRC64;
```

```
Query Match             2.3%;
Best Local Similarity  19.3%;         Score 211;  DB 1;  Length 2358;
Matches  371;  Conservative 212;  Mismatches 700;  Indels 642;  Gaps  88;
```

```
Qy     27 DAPTTRTTCTNNSPKQFDASRSF----------------------NEVELTKVETKG-GT    63
             ::::::::: ::|
Db    534 DSSVSLSTQTLNADSHSTATLTFIAHDAAGNPVVGLVLSTRHEGVQDITLSDWKDNGDGS   593

Qy     64 FASVSIPGAF------------------PTGEVGSPEVPAVRKLIAV----------PVG    95
          ||: |                             :|||::  : :
Db    594 YTQILTTGAMSGTLTLMPQLNGVDAAKAPAVVNIISVSSRTHSSIKIDKDRYLSGNPIE   653

Qy     96 ATPVVR---VKSFTEQVYSLNQYGSEKLMPHQPSMSKSDDPEKVPFVYNA-AAYAR--   147
              |||:     :: :  || || :::
Db    654 VTVELRDENDKPVKEQKQQLN-NAVSIDNVKPGVT-TDWKETADGVYKATYTAYTKGSG   710

Qy    148 -------------------KGFVGQELTQVEMLGTMR---------GVRIAALTIN    175
                                 |:::                    ||
Db    711 LTAKLLMQNWNNEDLHTAGFIIDANPQSAKIATLSASNNGVLANENAANTVSVNVADEGSN   770

Qy    176 PVQYDVV---------ANQLKVRNNIEIEVSFQGADEVATQRLIYDASFSPYFETAYKQL--   225
          |::                                                    : :|
Db    771 PINDHTVTFAVLSGGSATSFMNQNTAKTDVNGLATFDLKSSKQEDNTVEVTLENGVKQTLI   830

Qy    226 --FNRDVYTDHGDLYNTPVRMLVVAGAKFKEALKPWLTWKAQKGFYL-DVHYT---DEAE   279
                                                              :: |
Db    831 VSFVGDSSTAQVDLQKSKNEVV----ADGNDSVIWTATVRDAKGNLLNDVMVTFNVNSAE   886

Qy    280 VGTTNASIKAFIHKKYNDGLAASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGD   339
             :  :                                 :
Db    887 AKLSQTEVNS--------HDGIATATL--------TSLKNGDY-RVTASVSSGSQANQQVN   930
```

```
RESULT 14
YEEJ_ECOLI
ID   YEEJ_ECOLI    STANDARD;    PRT;    2358 AA.
AC   P76347; P94750;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein yeeJ.
GN   YEEJ OR B1978.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251358; PubMed=9097040;
RA   Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA   Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA   Saito N., Sampei G., Seki Y., Sivasundaram S., Tagami H.,
RA   Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 40.1-50.0 min region on the linkage map.";
RL   DNA Res. 3:379-392(1996).
CC   -!- SIMILARITY: Contains 13 Big-1 domains.
```

```
Db   1270 LAGTFPNATWDPSAQQWTKID-NNTYSITLTLDEGTQIEYKYARGSWDKVEKDEYGNEFA 1328
              || :::  ||||   |||:
Qy   1204 ----------------------------------------SNV----VSSFSWNG--MA 1216
                                                      ||:  :  :||   :
Db   1329 SNRKVTIVNQGNNEMTINDTVRWRDIPIFIYSPSSNMTVDSNISTMEVKGNTYKGAKVT 1388
          ||:  :  ::|
Qy   1217 LNPDNY---LISKDVT---GATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFT---- 1264
          ::: || || |   ||   || ||
Db   1389 INGDSFVQDKNGVFTKDVSLNYGVNKIKIHVEPNDG--------SVYGNDQGRITELT 1438
                                              |::
Qy   1265 -----VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCS 1318
               :::  : |: ||:  : : :
Db   1439 KDIEIDVIRQENNSGSGTGNN----NTSTSGSNSSS 1470
          |:  :: : |
Qy   1319 DLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVK 1378
              |   :
Db   1471 --------TGSGSTGSTSITSNISN-------TSNISNTIGVITKNGN-VITLT 1508
          :||||||:|||||:|: |:
Qy   1379 YTAGVSPKKCVDVTVNSTQFNPVQNLTA- 1420
          |||| : |||:|||    |:||:
Db   1421 APASKRAEVLNED-FENGIPASWKTIDADGDGN--NWTTPPGGGSFPAGHNSAICVSSA 1477
             ::: ::|: |   ::
Qy   1509 LDAG---KAKDLIVNSKDKKVVFDITTIGEGQQKVVQISKDILDTSAANGKDIVIKSDN 1564
          || |      :  :  |||
Db   1478 SHI--NFEGPQNPDNYLV------TPELSLPGG-----GTLTFWVCAQDANYASEHYAV 1523
             :|:|| :   |||:
Qy   1421 APASKRAEVLNED-FENGIPASWKTIDADGDGN--IKDNGKPNVTNYVSLSNVVDITISGISGNVTIAKP 1621
            ||: |   |::|  :  | ||  |||
Db   1565 ASIALTKDALNQNQIQNGVNVS---IKDNGKPNVTNYVSLSNVVDITISGISGNVTIAKP 1621
          |||:|:||: | :|  :
Qy   1478 SHI--NFEGPQNPDNYLV------TPELSLPGG-----GTLTFWVCAQDANYASEHYAV 1523
          |     ||
Db   1622 VEVTLNISKANDPRKVAVYYYNPTTNQWEYVGGKVDASSGTITF-----NATHFSQYAAF 1676
          ||| :      | | |||
Qy   1524 YASSTGND-ASNFANALLEEVLTAKIVV 1550
          :|| ||  |:|:|||
Db   1677 EYDKIFNDIKDNWAKDVI-EVLASRHIV 1703
```

```
Qy   1695  NPVEGIKVNF------RGTSVTLSSTSVETDDRGFAEILVTSTEVGLKTVSASILADKPT 1747
Qy   1428  EVLNEDFENGIPASWKTIDADGDGNNWTTT--PPPGGSSFAGHNSAICVSSASHIN--FE 1483
Db   1748  EVIS--------RLLNASADVNSATTTSLEIPEGQVMVAQDVAV---KAHVNDQFG 1792
Qy   1484  GP------QNPDNYLV-------------------TPE------LSLPGGGTLTFW 1508
Db   1793  NPVAHQPVTFSAEPSSQMIISQNTVSTNTCQGVAEVTMTPERNGSYMVKASLPNGASLEKQ 1852
Qy   1509  VCAQDANY-----ASEHYAVTASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGT 1564
Db   1853  LEAIDEKLTLTASSPLIGVYAP----------TGATLTATLTSANGTPVEGQ 1894
Qy   1565  WYQKTVQLPAGTKY---VAFRHFGCTDFFWINLDDVVITSGNAPSYTYTI-YRNNTQIA 1619
Db   1895  VINFSV-TPEGATLSGGKVRTNSSGQA----------PVVLTSNKVGTYTVTASFHNGVTIQ 1945
Qy   1620  SGVTETTYR---------------DP----DLAT-------------GFYTGVKVVYPN 1647
Db   1946  ---TQTTVKVTGNSTAHVASFIADPSTIAATNTDLSTLKATVEDGSGNLIEGLTVYF-- 2000
Qy   1648  GESAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGBEAMIYDMNGRRLIAAGRNTVVY 1707
Db   2001  --ALKSGSATLTSLTAVTDQNGIATTSVKGAMTGSVTVSAV--------TTAGGMQTVDI 2050
Qy   1708  TAQGG 1712
Db   2051  TLVAG 2055
```

```
RESULT 15
OMPB_RICCN
ID   OMPB_RICCN     STANDARD;     PRT;   1655 AA.
AC   Q9KKA3; Q9KK98; Q9XC45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB OR RC1085.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Malish 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL   Science 293:2093-2098(2001).
RN   [2]
RP   SEQUENCE OF 33-1649 FROM N.A.
RC   STRAIN=Indian tick typhus, and Malish 7;
RX   MEDLINE=20393643; PubMed=10939649;
RA   Roux V., Raoult D.;
RT   "Phylogenetic analysis of members of the genus Rickettsia using the
RT   gene coding the outer-membrane protein rOmpB (ompB).";
RL   Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN   [3]
RP   SEQUENCE OF 353-1655 FROM N.A.
RC   STRAIN=Malish 7;
RA   Stenos J., Walker D.;
RT   "The rickettsial outer membrane protein A and B genes of Rickettsia
RT   australis, the most divergent rickettsia of the spotted fever group.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
```

```
Qy    340  YFPEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQP- 398
Db    931  FIGDQSTRALTLSVPSGDITVTNTAPQYMTATLQDKN------------------GNPL 971
Qy    399  ----------TIKYGMQYYYNQEHGYTDVNYLKAPYTGCY--SHI---------NTGVS 437
Db    972  KDKEITFSVPNDVASKFSIS--NGGKGMTDSNGVAIASLTGTLAGTHMIMARLANSNVS 1028
Qy    438  FANYTAHGSETAWADPLLITSQLKALTNKDKYFLAIGNCCITRAQFDYVQPCFGEVITRVK 497
Db   1029  DAQPMTFVADKDRAVVVLQTSKAEIIGN------GVDETTLTA------------TVK 1068
Qy    498  EKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSINTVNSIM 557
Db   1069  DPSNHPVAGITVNFTMPQD----VAANFT--------------LENN----------- 1097
Qy    558  WAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNTYTLPASLPQNQAS 617
Db   1098  --------GIAITQANGEAHVTLKG-------------KKAGTHTVTATLGNNNTS 1132
Qy    618  YS-------IQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN 670
Db   1133  DSQPVTFVADKASAQVVLQISKDEIT-GNGVDSATLTATVK-----------DQFDNEVN 1180
Qy    671  YLPVIKQIVQGEPSPYQPVSNLTATTQGQKVTLK-----WEAPSAKKAEGSREVKRIGDGLF 727
Db   1181  NLPV------TFSSASSGLLTPGVSNTNESGIAQATLAGVAFGEKTV----------- 1222
Qy    728  VTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASSNL 787
Db   1223  ------------TASLANNGASDNKTVHFIGDTAAAKIIELAPVPDSIIAGT------- 1262
Qy    788  YSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIIAGDGNQ 847
Db   1263  ------------------PQNSSGSVITATVV-------------------DNNGF 1281
Qy    848  PARYDDFTFEAGKKYTFTMR--RAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTAT 905
Db   1282  PVK------GVTVNFTSNAATAEMTNGGQAVTNEQGKATVTYTNTR--SSIESGARPD 1331
Qy    906  TFE---EDGVAAGNHEYCVEVKTAGVSPKVCKDVTVEGSNEFAPVQNLTGS-SVGQKVT 961
Db   1332  TVEASLENGSS---------TLSTSINVNADAS--TAHLTLLQALFDTVSAGETTS 1376
Qy    962  LKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYN 1021
Db   1377  LYIE---------------VKDNYGNGVPQQEVTLSV-SPSEGVTPSN-NAIYTTN 1415
Qy   1022  SNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWCVAQDANYASEHYAVYASST 1081
Db   1416  HDGNFYA-SFTATKAGV----YQLTATLE--NGDSMQQTV-TYVPNVANAEITLAASKD 1466
Qy   1082  GNDASNFTNALLEETIITAKGVSRVSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMF 1141
Db   1467  PVLIADN--NDLTTLTATVADTEG-----NAIANT--EVTFTLPEDVK-----------ANFT 1508
Qy   1142  YIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDQGWLCLSSGQL------- 1195
Db   1509  LSDGGKVITDAEGKAKVTLKGTKAGAH---------TVTASMTGG----KSEQLVVNFIA 1555
Qy   1196  DWLTAHGGSNVVSSFSWNGMALNPDNVLISKDVTGATKVKYYYAVNDGFGPGDHYAVMISK 1255
Db   1556  DTLTAQVNLNVTE----------DNFTANN--VGMTRLQATVTDGNGNPLANEAVTFTL 1602
Qy   1256  TGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTK------Y 1309
Db   1603  PADVSASFT---------LGQGGSAI--TDINGKAEVTL-----------SGTIKSGTYPVT 1641
Qy   1310  VAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATG 1369
Db   1642  VSVNNYGVSDTKQVTL---IADAGTAKLIASLT-SVYSFVVSSTTEGATMTAGVTD--ANG 1694
Qy   1370  NHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP--ASKRA 1427
```

```
CC      STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC      VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC      similarity).
CC      -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC          (By similarity).
CC      -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC          layer with hexagonal symmetry (By similarity).
CC      -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC      ----------------------------------------------------------------------
```
```
DR      EMBL; AE008659; AAL03623.1; -.
DR      EMBL; AF123721; AAF34124.1; -.
DR      EMBL; AF123726; AAF34129.1; -.
DR      EMBL; AF149110; AAD39533.1; -.
DR      PIR; E97835; E97835.
DR      InterPro; IPR006315; Autotransport.
DR      InterPro; IPR005546; Autotransporter.
DR      Pfam; PF03797; Autotransporter, 1.
DR      TIGRFAMs; TIGR01414; autotrans_barl, 2.
KW      Antigen; S-layer; Cell wall; Complete proteome.
FT      CHAIN         1    1334       120 kDa SURFACE-EXPOSED PROTEIN.
FT      CHAIN      1335    1655       32 kDa BETA PEPTIDE.
FT      VARIANT      61      61       P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT      VARIANT      75      75       G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT      VARIANT      78      78       K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT      VARIANT     251     251       V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT      VARIANT     413     413       N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT      VARIANT     959     959       I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT      VARIANT     988     988       A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT      VARIANT    1139    1139       R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT      CONFLICT    353     354       KD -> GH (IN REF. 3).
FT      CONFLICT    776     776       F -> S (IN REF. 3).
FT      CONFLICT   1159    1159       E -> D (IN REF. 3).
FT      CONFLICT   1177    1177       G -> D (IN REF. 3).
FT      CONFLICT   1492    1492       H -> R (IN REF. 3).
SQ      SEQUENCE   1655 AA;  168342 MW;  E49E19377D5FCE37 CRC64;

Query Match             2.3%;  Score 209.5; DB 1; Length 1655;
Best Local Similarity  19.2%;  Pred. No. 0.0019;
Matches  277;  Conservative 153;  Mismatches 501;  Indels 511;  Gaps 67;

Qy    499 KGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMG-------------SYDA-T 544
Db    383 KTAASKVAITQNSNFGTTDFGNLAAQIIVPNTMTLNGNFTGDASNPGNTAGVITFPDANGT 442

Qy    545 FLEDS-------YNTVNSIMWAG---NLAATHA-------GNIGNITHIGAHYWEAXHVL 588
Db    443 LASASADANVAVTNNITAIEASGAGVVQLSGTHAAELRLGNAGSV---------FKL 490

Qy    589 GDGSVMPYRAMPKTNYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVA 648
Db    491 ADGTVINGKV-----------------NQTALVGGALAAGTITLDGSATITG-DIGNAGGAA 534

Qy    649 TV-------SMTKQIT-----------ENGNYDVVITRSNYLPVIKQIQVGEP 683
Db    535 ALQGITLANDATKTLTLGGANIIGANGGTINFQANGGTIKLTSTQNN---IVVDFDLAIA 591

Qy    684 SPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPANDVRANEAKV 743
Db    592 TDQTGVVDASSLTNAQLTIINGKIGTV--GANNKTLGQFNIGSSKTVLSDGDVAINEL-- 647

Qy    744 VLAADNVWGDNTGYQFLLDADHNT------------PGSVIPATGPLFTGT----A 783
Db    648 ------VIGNNGAVQFA----HNTYLITRTNAAGQGKIIFNPVVNNNTTLATGTNLGSA 697

Qy    784 SSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEP--------AS 835

Db    698 TNPLAEINFGSKGAANVDTVLN-----------VGKGV--------NLYATNITTTDANVGSFIFNAG 746
Qy    836 GKWWIAGD-GGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTMEVEDDSPASYTYT--- 890
Db    747 GTNIVSGTVGGQQGNKFNTVALDNGTTVKF-LGNATFNGNTTIAANSLIQIGNYTADFV 805
Qy    891 VYRDGTKIKEGLTATTEEDGVAAGNHEYCVEVKYTAGVSP-KVCKDVTVEGS-----NE 944
Db    806 ASADGTGIVE----------FVNTGP------ITVTLNKQAAPVNALKQTVSGPGNVVINE 851
Qy    945 FAPVQNLTG----SSVGQKVTLKWDAPNGTP-NPNENPNP-----NPGTTLSES 988
Db    852 IGNAGNVHGAVTDTIAFENSLGAVVFL----PRGIPFNDAGNRIPLTIKSTVGNKTATG 907
Qy    989 FB--NGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLI 1046
Db    908 FDVPSVILGVDSVIADGQVIG-DQNNIVGLGLGSDNDIIVNATTLYAGIGTINNNQGTV 966
Qy   1047 TPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF------------T 1089
Db    967 TLSGGIPN------------------------TPGTVYGLGTGIGASKFKQVTFTTDYNNLGNIIAT 1009
Qy   1090 NALLEE--TITAKGVR----SPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143
Db   1010 NATINDGVTVTTGIAGIGFDGKITLGSVNGNGNVRFVD---------------- 1048
Qy   1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGG 1203
Db   1049 --------------------------------GILSHSTSMIGTTKANNG 1066
Qy   1204 SNVVSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAG-- 1261
Db   1067 TVTYLGNAFVGNIGDSDTPVASVRFTGSDG--------GAGLQGNIYSQVIDFGTYNLGIS 1119
Qy   1262 DFTVVFEETPNGIN-KGGARFGLSTRANGAKPQSVW----IERTVDLPAGTKYVAFRHY 1315
Db   1120 NSNVILGGGTAINGKINLRKTNTLTFASGT--STWGNNTSIETTLTLANG------ 1167
Qy   1316 NCSDLNYILLDDIQFTMGGSPTPT----------DYTYTVYRDGTKIKEGLTETTFE 1362
Db   1168 NIG--NIVILEGAQVNATTTGTTTIKYQDNANANFSGTQTYTLIQGGARFNGTLGGGPNF- 1224
Qy   1363 EDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422
Db   1225 ---VVTGSNRF---VNY-GLIRAANQDYVI-TRTNNAENVVTN--------------- 1259
Qy   1423 ASKRAEVLNEDFENGIPASWKTIDADGDGNWTTTPPGGSSFPAGHNSAICVSSASHINF 1482
Db   1260 ------DIANSSF-GGAP--------GVGQN-VTTFVNATNTPAAYNNLLAKNSANSANF 1303
Qy   1483 EGPQNPD---------------------------NYLVTP---ELSLPGGGTLTFWVCAQ 1512
Db   1304 VGAIVTDTSAAITNAQLDVAKDIQAQLGNRLGALRYLGTPETAEMAGPEAGAIPAAVAAG 1363
Qy   1513 DANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQL 1572
Db   1364 DEAVDNVAYGIWAKPFVTDAHQ---------------SKKGGLAGYKAKTT--------- 1399
Qy   1573 PAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQI--ASGVTETTYRDP 1630
Db   1400 ---------GVV---IGLD---------TLANDNLMIGAAIGTKTDIKHQ 1429
Qy   1631 DLATG--------FYTYG---VKVVYPNGESAIETATLNITSLADVTAQKPYTLTVG 1677
Db   1430 DYKKGDKTDVNGFSFSLYGAQQLVKNFFAQG-SAI--FSLN------------- 1467
Qy   1678 KTITVTCQGEAMIYDMNG---RRLAAGRN----------TVVYTAQGHYAVMVVVDGKSY 1725
Db   1468 --QVKNKSQRYFFDANGNMSKQIAAGHYDNMTFGGNLTVGYDNAMQGVLVTPMAGLSY 1524
Qy   1726 VE 1727
```

us-08-353-485-10.rsp

Db   1525 LK 1526

Search completed: May 18, 2004, 11:33:32
Job time : 35.7568 secs

```
            GenCore version 5.1.6
      Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:       May 18, 2004, 11:28:29 ; Search time 69.1557 Seconds
                                       (without alignments)
                                       7902.136 Million cell updates/sec

Title:        US-08-353-485-10
Perfect score: 9179
Sequence:     1 MRKLLLIAASLLGVGLYAQ.........HYAVMVVVDGKSVVEKLAVK 1732

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:     1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database  :   SPTREMBL_25:*
              1: sp_archea:*
              2: sp_bacteria:*
              3: sp_fungi:*
              4: sp_human:*
              5: sp_invertebrate:*
              6: sp_mammal:*
              7: sp_mhc:*
              8: sp_organelle:*
              9: sp_phage:*
             10: sp_plant:*
             11: sp_rodent:*
             12: sp_virus:*
             13: sp_vertebrate:*
             14: sp_unclassified:*
             15: sp_rvirus:*
             16: sp_bacteriap:*
             17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	9179	100.0	1732	2	Q51817	porphyromon
2	9167	99.9	1732	2	O07442	porphyromon
3	9158	99.8	1732	2	O52050	porphyromon
4	8029.5	87.5	1723	2	P72194	porphyromon
5	7774.5	84.7	1723	2	P72197	porphyromon
6	6486	70.7	1223	2	Q9ZNB5	porphyromon
7	6156.5	67.1	1358	2	P96967	porphyromon
8	4272	46.5	1706	2	Q51839	porphyromon
9	4270	46.5	1704	2	Q51816	porphyromon
10	4256	46.4	1706	2	Q51838	porphyromon
11	4249.5	46.3	1097	2	P72196	porphyromon
12	4245	46.2	1687	2	Q9R9B7	porphyromon
13	1854	20.2	364	2	P96966	porphyromon
14	980.5	10.7	925	2	Q9F4J0	porphyromon
15	722	7.9	312	2	Q9KIB3	porphyromon
16	547	6.0	736	2	Q51844	porphyromon

17	541	5.9	736	2	O33441	porphyromon
18	354.5	3.9	293	2	Q9XBU9	porphyromon
19	341.5	3.7	422	2	Q51818	porphyromon
20	322	3.5	5291	16	Q8X2T1	escherichia
21	315	3.4	5188	16	Q8X4H5	escherichia
22	313	3.4	1993	16	Q92FG5	listeria in
23	309	3.4	1744	16	Q82YW8	enterococcu
24	306.5	3.3	2219	16	Q88W19	lactobacill
25	303	3.3	4210	16	Q89J13	bradyrhizob
26	300.5	3.3	2468	16	Q9I2M3	pseudomonas
27	300	3.3	8173	16	Q7UDU8	rhodopirell
28	295.5	3.2	1817	17	Q8TI59	methanosarc
29	295.5	3.2	3346	16	Q7WN54	bordetela
30	295.5	3.2	4199	16	P74440	synechocyst
31	295	3.2	1975	2	Q8GJ03	enterococcu
32	294	3.2	2522	16	Q8EKA6	shewanella
33	291	3.2	2523	17	Q8TJS8	methanosarc
34	289	3.1	3988	17	Q8TPZ1	methanosarc
35	285	3.1	1861	16	Q89L34	bradyrhizob
36	284.5	3.1	1752	2	Q9AE52	ruminococcu
37	283	3.1	2566	17	Q8TSE7	methanosarc
38	280.5	3.1	3029	16	Q55582	synechocyst
39	280	3.1	2016	17	Q8TJS9	methanosarc
40	280	3.1	2230	16	Q7U7J7	synechococc
41	279.5	3.0	8682	16	Q88RG2	pseudomonas
42	278.5	3.0	1541	16	Q7UQJ9	rhodopirell
43	278.5	3.0	6310	16	Q88PP2	pseudomonas
44	276.5	3.0	2044	16	Q8Y8Q4	listeria mo
45	274.5	3.0	1673	1	Q977V5	methanosarc

ALIGNMENTS

```
RESULT 1
Q51817
ID   Q51817               PRELIMINARY;    PRT;  1732 AA.
AC   Q51817;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Porphypain.
GN   PRTP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W12;
RX   MEDLINE=96213011; PubMed=8631659;
RA   Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA   Progulske-Fox A., Lantz M.S.;
RT   "Analysis of the prtP gene encoding porphypain, a cysteine proteinase
RT   of Porphyromonas gingivalis.";
RL   J. Bacteriol. 178:2734-2741(1996).
DR   EMBL; U42210; AAB06565.1; -.
DR   PIR; T30836; T30836.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase C25 C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE  1732 AA;  187875 MW;  65427lDBEF7BCAE4 CRC64;
```

```
Query Match       100.0%;  Score 9179;  DB 2;  Length 1732;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 1732;   Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTCTNNSPKQFDASFSFNEVELTKVETK 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTCTNNSPKQFDASFSFNEVELTKVETK 60

Qy    61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVRVKSFTEQVYSLNQYGSEKL 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVRVKSFTEQVYSLNQYGSEKL 120

Qy   121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGPVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGPVGQELTQVEMLGTMRGVRIAALTINPVQYD 180

Qy   181 VVANQLKVRNNIEIBVSFQGADEVATORLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 VVANQLKVRNNIEIBVSFQGADEVATORLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240

Qy   241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300

Qy   301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360

Qy   361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNY 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNY 420

Qy   421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480

Qy   481 QFDYVVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSGANAVFGVQPTFEGTSMGS 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 QFDYVVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSGANAVFGVQPTFEGTSMGS 540

Qy   541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP 600

Qy   601 KTNYTLPASLPQNCASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 KTNYTLPASLPQNCASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660

Qy   661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720

Qy   721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Qy   781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVDYCIITNPEPASGKMWI 840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVDYCIITNPEPASGKMWI 840

Qy   841 AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900

Qy   901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEEAPVQNLTGSSVGQKV 960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEEAPVQNLTGSSVGQKV 960

Qy   961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020

Qy  1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDPANYASEHYAVYASS 1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDPANYASEHYAVYASS 1080

Qy  1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140

Qy  1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200

Qy  1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260

Qy  1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320

Qy  1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380

Qy  1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440

Qy  1441 SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP 1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP 1500

Qy  1501 GGGTLTFWVCAQDADANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 GGGTLTFWVCAQDADANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560

Qy  1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTTTIYRNNTQIAS 1620
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTTTIYRNNTQIAS 1620

Qy  1621 GVTETTYRDPDLATGFYTGYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 GVTETTYRDPDLATGFYTGYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680

Qy  1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
```

RESULT 2
O07442
ID   O07442   PRELIMINARY;   PRT;   1732 AA.
AC   O07442;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine-specific cysteine proteinase.
GN   PRTK.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=99235907; PubMed=10219167;
RA   Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtK that encodes
RT   a lysine-specific cysteine proteinase and three sequence-related
RT   adhesins.";
RL   Oral Microbiol. Immunol. 14:92-97(1999).
DR   EMBL; U75366; AAB60809.1; -.
DR   MEROPS; C25.002; -.

```
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE  1732 AA;  187914 MW;  45D5B9137739I703 CRC64;

Query Match     99.9%;   Score 9167;   DB 2;   Length 1732;
Best Local Similarity  99.8%;   Pred. No. 0;
Matches 1729;   Conservative 3;   Mismatches 0;   Indels 0;   Gaps 0;

Qy     1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK   60
Db     1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK   60

Qy    61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSSKL  120
Db    61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSSKL  120

Qy   121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD  180
Db   121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD  180

Qy   181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYIDHGDLYNT  240
Db   181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT  240

Qy   241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA  300
Db   241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA  300

Qy   301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360
Db   301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360

Qy   361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNY   420
Db   361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNY   420

Qy   421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA  480
Db   421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA  480

Qy   481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS  540
Db   481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS  540

Qy   541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP   600
Db   541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP   600

Qy   601 KINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG  660
Db   601 KINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG  660

Qy   661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK  720
Db   661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK  720

Qy   721 RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGVYQFLLDADHNTFGSVIPATGPLFT  780
Db   721 RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGVYQFLLDADHNTFGSVIPATGPLFT  780

Qy   781 GTASSNLYSANFEYLIVPANADPVVTQNIIVTGGQEVVIPGGVYDYCITNPEPASGKMWI  840
Db   781 GTASSNLYSANFEYLIPANADPVVTQNIIVTGGQEVVIPGGVYDYCITNPEPASGKMWI   840

Qy   841 AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE   900
Db   841 AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE   900

Qy   901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVBGSNEFAPVQNLTGSSVGQKV  960
Db   901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVBGSNEFAPVQNLTGSSVGQKV  960

Qy   961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY  1020
Db   961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY  1020

Qy  1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080
Db  1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080

Qy  1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
Db  1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140

Qy  1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200
Db  1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200

Qy  1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGPPGDHYAVMISKTGTNA  1260
Db  1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGPPGDHYAVMISKTGTNA  1260

Qy  1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
Db  1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320

Qy  1321 NYILLDDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT  1380
Db  1321 NYILLDDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT  1380

Qy  1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
Db  1381 AGVSPKKCVNVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440

Qy  1441 SWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVSSASHINFEGQNPDNYLVTPELSLP  1500
Db  1441 SWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVSSASYINFEGQNPDNYLVTPELSLP  1500

Qy  1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
Db  1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560

Qy  1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIAS 1620
Db  1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIAS 1620

Qy  1621 GVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680
Db  1621 GVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680

Qy  1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK  1732
Db  1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK  1732

RESULT 3
052050
ID   052050        PRELIMINARY;        PRT;   1732 AA.
AC   052050;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine specific cysteine protease.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
```

```
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=98298016; PubMed=9632563;
RA   Lewis J.P., Macrina F.L.;
RT   "IS195, an insertion sequence-like element associated with protease
RT   genes in Porphyromonas gingivalis.";
RL   Infect. Immun. 66:3035-3042 (1998).
DR   EMBL; AF017059; AAC26523.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase C25.
DR   InterPro; IPR005536; Peptidase C25_C.
DR   Pfam; PF01364; Peptidase C25; 1.
DR   Pfam; PF03785; Peptidase C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE   1732 AA;  187931 MW;  B2337463D5CB5EA5 CRC64;

Query Match             99.8%;  Score 9158;  DB 2;  Length 1732;
Best Local Similarity   99.8%;  Pred. No. 0;
Matches  1729;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCNNSFKQFDASFSFNEVELTKVETK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCNNSFKQFDASFSFNEVELTKVETK 60

Qy         61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL 120

Qy        121 MPHQPSMSKSDDPEKVPFVVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MPHQPSMSKSDDPEKVPFVVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180

Qy        181 VVANQLKVRNNIEIBVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
              ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db        181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240

Qy        241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300

Qy        301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360

Qy        361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVSQPTIKYGMQYYYNQBHGYTDVYNY 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVSQPTIKYGMQYYYNQBHGYTDVYNY 420

Qy        421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480

Qy        481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYWSVGANAVFGVQPTFEGTSMGS 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYWSVGANAVFGVQPTFEGTSMGS 540

Qy        541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP 600

Qy        601 KTNTYTLPASLPQNQASYSIQASAGGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 KTNTYTLPASLPQNQASYSIQASAGGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660

Qy        661 NYDVVITRSNLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NYDVVITRSNLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720

Qy        721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Qy        781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVYDYCITNPEPASGKMWI 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVYDYCITNPEPASGKMWI 840

Qy        841 AGDGGNQOPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTYRDGTKIKE 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGDGGNQOPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTYRDGTKIKE 900

Qy        901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960

Qy        961 TLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY 1020

Qy       1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080

Qy       1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140

Qy       1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200

Qy       1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260

Qy       1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320

Qy       1321 NYILLDDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFBEDGVATGNHEYCVEVKYT 1380
              |  |||||||||||||||||||||||||| ||||||||||| | ||||||||||||||||
Db       1321 NVILLDDIQFTMGGSPTPTDYTYTVVRDGTNIKEGLTETTFBEYGVATGNHEYCVEVKYT 1380

Qy       1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440

Qy       1441 SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNVLVTPELSLP 1500
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 SWKTIDAHGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNVLVTPELSLP 1500

Qy       1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEALRGTR 1560
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEALRGTR 1560

Qy       1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYIYRNNTQIAS 1620
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYIYRNNTQIAS 1620

Qy       1621 GVTETTYRDPDLIATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GVTETTYRDPDLIATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680

Qy       1681 TVTCQGBAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVDGKSYVEKLAVK 1732
              |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVDGKSYVEKLAVK 1732

RESULT 4
P72194
ID    P72194;            PRELIMINARY;        PRT;   1723 AA.
AC    P72194;
DT    01-FEB-1997 (TrEMBLrel. 02, Created)
DT    01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE    Lys-gingipain.
GN    KGP.
OS    Porphyromonas gingivalis (Bacteroides gingivalis).
OC    Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC    Porphyromonadaceae; Porphyromonas.
OX    NCBI_TaxID=837;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=381;
RX    MEDLINE=97044756; PubMed=8889827;
RA    Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
RT    "Cloning and sequencing of the gene encoding a novel lysine-specific
RT    cysteine proteinase (Lys-gingipain) in Porphyromonas gingivalis:
RT    structural relationship with the arginine-specific cysteine proteinase
RT    (Arg-gingipain).";
RL    J. Biochem. 120:398-406(1996).
DR    EMBL; D83258; BAA11870.1; -.
DR    MEROPS; C25.002; -.
DR    GO; GO:0005524; F:ATP binding; IEA.
DR    GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR    GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR    GO; GO:0006310; P:DNA recombination; IEA.
DR    GO; GO:0006281; P:DNA repair; IEA.
DR    GO; GO:0006260; P:DNA replication; IEA.
DR    GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR    InterPro; IPR000977; DNA ligase.
DR    InterPro; IPR001769; Peptidase_C25.
DR    InterPro; IPR005536; Peptidase_C25_C.
DR    Pfam; PF01364; Peptidase_C25; 1.
DR    Pfam; PF03785; Peptidase_C25_C; 1.
DR    PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ    SEQUENCE   1723 AA;  187261 MW;  5628963D251493BB CRC64;

Query Match             87.5%;
Best Local Similarity   88.2%;
Matches 1537;  Conservative  60;  Mismatches  116;  Indels  29;  Gaps  10;
                        Score 8029.5;  DB 2;  Length 1723;
                        Pred. No. 0;

Qy    1 MRKLLLLIAASLLGVGLYAQSAIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRKLLLLIAASLLGVGLYAQSAIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK 60

Qy   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL 120

Qy  121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180

Qy  181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKLQLFNRDVYTDHGDLYNT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKLQLFNRDVYTDHGDLYNT 240

Qy  241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300

Qy  301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360

Qy  361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSCVGQOPTIKYGMQYYYNOEHGYTDVVNY 420
          ||||| ||||| ||||||  ::|| |||||  || |||   ||||||:   |||||  :|:
Db  361 IDKVLMYEKATMPDKSYLEKALLIAGADSYWNPKIGQQTIKYAVQYYNQDHGYTDVYSY 420

Qy  421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALINKDKYPLAIGNCCITA 480
          ::||||||||||||||||||||||||||||:||||||||:||:|||||::||||||||:|
Db  421 PKAPYTGCYSHLNTGVSFANYTAHGSETSWADPSLTATQVKALINKDKYFLAIGNCCVTA 480

Qy  481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS 540
          |||||||||||  ::||||||||||||||||||||||||||||||||||||||||||||
Db  481 QFDYPQPCFGEVMTRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS 540

Qy  541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP 600

Qy  601 KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db  601 KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVNMTKQITENG 660

Qy  661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720
          ||||||||||||||||||||||| ||||||||||||||||||||||||:||||||:||||
Db  661 NYDVVITRSNYLPVIKQIQAGEPSPYQPVSNLTATTQGQKVTLKWDAPSAKKAEASREVK 720

Qy  721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780
          ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db  721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Qy  781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWI 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWI 840

Qy  841 AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTTVYYRDGTKIKE 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db  841 AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTTVYYRDGTKIQE 900

Qy  901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKV 960

Qy  961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGHGWKPGNAPGIAGY 1020
          ||||||||||||||||||||| :|||||||||||||||||||||||||||||||||||
Db  961 TLKWDAPNGTPNPNPNPNPGT-TLSESFENGIPASWKTIDADGDHGHGWKPGNAPGIAGY 1019

Qy 1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1020 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1079

Qy 1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1080 TGNDASNFTNALLEETITAKGVRSPBAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1139

Qy 1141 FYIDLDEVEIKANGKRADFTETPESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1140 FYIDLDEVEIKANGKRADFTETPESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTA 1199

Qy 1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1200 HGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1259

Qy 1261 GDFTVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1260 GDFTVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1319

Qy 1321 NYILLDDIQFTMGGSPTPTDYTVYYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1320 NYILLDDIQFTMGGSPTPTDYTVYYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1379

Qy 1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFE-NGIP 1439
          |||||| :||||| ||::  ||||||  :|||||||    |||  ||   ||:|||:|
Db 1380 AGVSPKVCVNVTINPTQFNFVKNLKAQ--PDGGDVVLKWEAPSGKRGELLNEDFEGDAIP 1437

Qy 1440 ASWKTIDADGDGNWW--TTTPPGG------SSFAGHNSAICVSSASHINFEGPQNPDNYL 1492
          |:||| ||||||:|   ||:|||       :::| |:|  :|||:| :|:| ||::|:|:
Db 1438 TGWTALDADGDGNNWDITLNEFTRGERHVLSPLRASNVAISYSSLLQGQEYLPLTPNNFL 1497
```

```
Qy  1493  VTPELSLPGGGTLTFWVCAQD-ANYASEHYAVYASSTGNDASNFANALLEVLTAKTVVT  1551
          ||:|:    :|:|
Db  1498  ITPKVE--GAKKITYKVGSPGLPQWSHDHYALCISKSGTRAADF-----EVIFEETMTY  1549

Qy  1552  APEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVIT-SGNAPSYTYT  1610
          ||:|                |||||||||||:|| :     |||||:     ||:||||
Db  1550  TQGGANLTREK-------DLPAGTKYVAFRHYNCTDVLGIMIDDVVITGEGEGPSYTYT  1601

Qy  1611  IYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKP  1670
          :|:    :|: ||||||||       |:     |:|| :|: :|      ||||||||||
Db  1602  VYRDGTKIQEGLTETTYRDAGMSAQSHEYCVEVKXAAGVSPKVCVDYIPDGVADVTAQKP  1661

Qy  1671  YTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVYTAQGGHYAVMVVVDGKSYVEKLA  1730
          |||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db  1662  YTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVYTAQGGYYAVMVVVDGKSYVEKLA  1721

Qy  1731  VK  1732
          :|
Db  1722  IK  1723
```

RESULT 5
P72197

```
ID   P72197;      PRELIMINARY;      PRT;   1723 AA.
AC   P72197;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA   Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT   "Molecular cloning and characterization of Porphyromonas gingivalis
RT   Lys-gingipain.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U54691; AAA99810.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1723 AA;  186831 MW;  4508A7E50197CEBD CRC64;
```

```
Query Match             84.7%;  Score 7774.5;  DB 2;  Length 1723;
Best Local Similarity   84.9%;  Pred. No. 0;
Matches 1505;  Conservative  60;  Mismatches 118;  Indels 89;  Gaps 13;
```

```
Qy    1  MRKLILLIAASLLGVGLYAQSAKIKLDAPTRTTCTNNSFKQFDASFSFNEVELTKVETK    60
         ||||||||||||||||||| |||||||||||||||||||||||||||||||||| ||||
Db    1  MRKLILLIAASLLGVGLYAQNAKIKLDAPTRTTCTNNSFKQFDASFSFNEVELITKVETK   60

Qy   61  GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRKSFTEQVVSLNQYGSEKL   120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRKSFTEQVVSLNQYGSEKL   120

Qy  121  MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD  180
         ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db  121  MPHQPSMSKSDDPEKVPFAYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD  180

Qy  181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVTTDHGLYNT   240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVTTDHGLYNT   240

Qy  241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTNASIKAFIHKKYNDGLA   300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTNASIKAFIHKKYNDGLA   300

Qy  301  ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360

Qy  361  IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVQPTIKYGMQYYYNQEHGYTDVYNY   420
         |||||||||||||||||||||:||||||| :|| :||||||:|:||||||:|||||||:|
Db  361  IDKVLMYEKATMPDKSYLEKALLIAGADSYWNPKIQQTIKYAVQYYYNQDHGYTDVYSY   420

Qy  421  LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA  480
         :|||||||||||||||:||||||||||| |||| :| :|:||||||:|||||||||||:|
Db  421  PKAPYTGCYSHLNTGVGFANYTAHGSETSWADPSVTATQVKALTNKNKYFLAIGNCCVTA  480

Qy  481  QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS  540
         ||||.||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  QFDYPQPCFGEVMTRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS  540

Qy  541  YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDSVMPYYRAMP   600
         ||||||||||||||||||||||||||||:||||:|||||||||||||||||||||||||
Db  541  YDATFLEDSYNTVNSIMWAGNLAATHAENIGNVTHIGAHYWEAYHVLGDSVMPYYRAMP   600

Qy  601  KINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVNMTKQITENG  660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  KINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG  660

Qy  661  NYDVVITRSNYLPVIKQIQAGEBPSPYQPVSNLTATTQGQKVTLKWDAPSAKKAEGSREVK 720
         ||||||||||||||||||.|||||||||||||||||||||||||||:||||||||||||||
Db  661  NYDVVITRSNYLPVIKQIQVGEBPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720

Qy  721  RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQPLLDADHNTFGSVIPATGPLFT  780
         |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db  721  RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQPLLDADHNTFGSVIPATGPLFT  780

Qy  781  GTASSNLYSANFEYLIPANADPVVTQNIIVTGGQGEVVIPDGVYDYCITNPEPASGKMWI  840
         |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db  781  GTASSNLYSANFEYLVPANADPVVTQNIIVTGGQGEVVIPDGVYDYCITNPEPASGKMWI  840

Qy  841  AGDGGNQOPARYDDFTPEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVRDGTKIKE  900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841  AGDGGNQOPARYDDFTPEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVRDGTKIKE  900

Qy  901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKV  960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db  901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV  960

Qy  961  TLKWDAPNGTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY  1019
         |||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||
Db  961  TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY  1020

Qy 1021  NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDDANYASEHYAVYASS 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1020  NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDDANYASEHYAVYASS 1079

Qy 1081  TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM  1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1080  TGNDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM  1139

Qy 1141  FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTA  1200
         |||||||||||||||||||||||||:|||||:||||||||||||||||||||||||||
Db 1140  FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTA  1199

Qy 1201  HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA  1260
         ||| ||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db 1200  HGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA  1259
```

```
FT   NON_TER
SQ   SEQUENCE   1223 AA;   131542 MW;   00225CD2BA9F91B3 CRC64;

          Query Match        70.7%;   Score 6486;  DB 2;  Length 1223;
          Best Local Similarity   99.0%;   Pred. No. 0;
          Matches 1211;   Conservative   9;   Mismatches   3;   Indels   0;   Gaps   0;

Qy     510  NSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGN  569
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1  NSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGN  60

Qy     570  IGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNYTLPASLPQNQASYSIQASAGSYVA  629
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61  IGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKTNYTLPASLPQNQASYSIQASAGSYVA  120

Qy     630  ISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIKQIVQGEPSPYQPV  689
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121  ISKDGVLYGTGVANASGVATVNMTKQITENGNYDVVITRSNYLPVIKQIQAGEPSPYQPV  180

Qy     690  SNLTATTQGQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAADN  749
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181  SNLTATTQGQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAADN  240

Qy     750  VWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLVPANADPVTTQNI  809
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241  VWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLIPANADPVTTQNI  300

Qy     810  IVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFTMRRA  869
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301  IVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFTMRRA  360

Qy     870  GMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGV  929
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361  GMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGV  420

Qy     930  SPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPNPGTTLSESF  989
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421  SPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPNPGTTLSESF  480

Qy     990  ENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPA  1049
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481  ENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPA  540

Qy    1050  LDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPKAIR  1109
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541  LDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPEAIR  600

Qy    1110  GRIQGTWRQKTVDLPAGTKYVAFRHYQSTMFYIDLDEVRIKANGKRADFTETFESSTHG  1169
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601  GRIQGTWRQKTVDLPAGTKYVAFRHYQSTMFYIDLDEVRIKANGKRADFTETFRSSTHG  660

Qy    1170  EAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVT  1229
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661  EAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVT  720

Qy    1230  GATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANG  1289
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721  GATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEANG  780

Qy    1290  AKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDTYTVYVYRDG  1349
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781  AKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDTYTVYVYRDG  840

Qy    1350  TKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQA  1409
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841  TKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTINPTQFNPVQNLTAEQA  900

Qy    1410  PNSMDAILKWNAPASKRAEVLNEDFENGIPASWKTIDADGDGNNWTTTPPGGSSFAGHN  1469
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901  PNSMDAILKWNAPASKRAEVLNEDFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHN  960

Qy    1470  SAICVSSASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVVASSTG  1529
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    1261  GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL  1320
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1260  GDFTVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL  1319

Qy    1321  NYILLDDIQFTMGGSPTPDTYTVYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT  1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1320  NYILLDDIQFTMGGSPTPDTYTVYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT  1379

Qy    1381  AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE------------  1428
            |||:||:|:||:|||||||||||||||:|:||||||||||||:||||||
Db    1380  AGVSPKKCVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKAEGSREVKRIGDGL  1437

Qy    1429  ---------------VLNEDFENGIPASWKTIDADGDGNNWTTTPPGGSSFAGHNSA  1471
                           :|:||||||||||||||||||||||||||||||||:|||||:
Db    1438  FVTIEPANDVRANEAKVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASS  1496

Qy    1472  ICVSSASHINFEG-----PQNPD------NYLVT--PELSLPGGGTLTFWVCAQDANYASEHY  1521
            ||:|||:::||:          :       :||:       :::||||||
Db    1497  NLYSA----NFEYLIPANADPVTTQNIIVTGQGEVVIPGG---VYDYCITNPEPASGKM  1549

Qy    1522  AVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-A  1580
            ::|||||||||||: |||:|::::  :: ::|:|:::  :|:|:|     ::||||||:|
Db    1550  WI-AGDGDGNQPARYDDFTFE--------------------------------AGKKYTFT  1576

Qy    1581  FRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPLATGFYTYG  1640
            |:|:|  ::   ::|::::::|:|:|:|:|:||:::|:|:|||||||||:::|:|:||
Db    1577  MRRAGMGDG-----TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYC  1631

Qy    1641  VKVVYPNGESAIETATLNITSLADVTAQKPYTLLTVVGKTITVTCQGEAMIYDMNGRRLAA  1700
            ||:|  |||:|:|::|:::|::|:|:|:||||:|:||||||:|:|:|||||||||||||||
Db    1632  VEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLLTVVGKTITVTCQGEAMIYDMNGRRLAA  1691

Qy    1701  GRNTVVYTAQGHYAVMVVVDGKSYVEKLAVK  1732
            |||||||||||:|||||||||||||||||||
Db    1692  GRNTVVYTAQGGYAVMVVVDGKSYVEKLAVK  1723
```

RESULT 6

```
Q9ZNB5
ID   Q9ZNB5      PRELIMINARY;      PRT;   1223 AA.
AC   Q9ZNB5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   130k-HMGD (Fragment)
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=99143166; PubMed=9988746;
RA   Shibata Y., Hayakawa M., Takiguchi H., Shiroza T., Abiko Y.;
RT   "Determination and characterization of the hemagglutinin-associated
RT   short motifs found in Porphyromonas gingivalis multiple gene
RT   products.";
RL   J. Biol. Chem. 274:5012-5020(1999).
DR   EMBL; AB019363; BAA34341.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005524; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25, 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
```

```
Db   961  SAICVSSASYINFEGQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1530  NDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDF 1589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  NDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDF 1080
Qy  1590  FWINLDDVVTSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYCVKVVYPNGE 1649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  FWINLDDVVTSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYCVKVVYPNGE 1140
Qy  1650  SAIETATLNITSLADVTAQKPYTLTVVGKTITVVCQGEAMIYDMNGRRLAAGRNTVVYTA 1709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  SAIETATLNITSLADVTAQKPYTLTVVGKTITVVCQGEAMIYDMNGRRLAAGRNTVVYTA 1200
Qy  1710  QGGHYAVMVVDGKSYVEKLAVK 1732
          ||||||||||||||||||||||
Db  1201  QGGHYAVMVVDGKSYVEKLAVK 1223
```

RESULT 7
P96967
ID   P96967   PRELIMINARY;   PRT;   1358 AA.
AC   P96967;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin.
GN   HAGD.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Lepine G., Whitlock J., Wojciechowski L., Progulske-Fox A.;
RT   "Cloning, sequencing and characterization of hagD, a member of the
RT   HArep multigene family in Porphyromonas gingivalis.";
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U68468; AAB49691.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   SMART; SM00060; FN3; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE 1358 AA; 147102 MW; 47FCA0B25B06DED8 CRC64;

```
Qy   Query Match      67.1%;   Score 6156.5;  DB 2;   Length 1358;
     Best Local Similarity   85.0%;  Pred. No. 2e-310;
     Matches 1171;   Conservative  59;  Mismatches 118;  Indels  29;  Gaps  10;

Qy   366  MYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQBEHGYTDYNYLKAPY 425
          ||||||||||||||| :||||||| |||| |:||||: |||||| :|||:||| |:
Db     1  MYEKATMPDKSYLEKALLIAGADSYWNPKIGQQTIKYAVQYYNQDHGYTDVYSYPKAPY 60
Qy   426  TGCYSHLNTGVSFANYTAHGSETAWADPLLTSQLXALTNKDKYFLAIGNCCITAQFDYV 485
          |||||||||||:||||:||||||||||| :|::|:||||||||||||||||:||||||
Db    61  TGCYSHLNTGVGFANYTAHGSETSWADPSLTATQVKALTNKDKYFLAIGNCCVTAQFDYP 120
Qy   486  QPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYWSVGANAVFGVQPTFEGTSMGSYDATF 545
          ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  QPCFGEVMTRVKEKGAYAYIGSSPNSYWGEDYWSVGANAVFGVQPTFEGTSMGSYDATF 180
```

```
Qy   546  LEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMPKTNTY 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  LEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMPKTNTY 240
Qy   606  TLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVV 665
          |||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db   241  TLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVNMTKQITENGNYDVV 300
Qy   666  ITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVKRIGDG 725
          |||||||||||:|| ||||||||||||||||||||||||||:|||||||:|||||||||
Db   301  ITRSNYLPVIKQIQAGEPSPYQPVSNLTATTQGQKVTLKWDAPSAKKAEASREVKRIGDG 360
Qy   726  LFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASS 785
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  LFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASS 420
Qy   786  NLYSANFEYLVPANADPVVTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGG 845
          ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||:
Db   421  NLYSANFEYLIPANADPVVTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGD 480
Qy   846  NQPARYDDFTEAGKKYTFMRRAGMGDGTMEVEDDSPASYTYTVVRDGTKIKEGLTAT 905
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  NQPARYDDFTEAGKKYTFMRRAGMGDGTMEVEDDSPASYTYTVVRDGTKIKEGLTAT 540
Qy   906  TFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWD 965
          ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db   541  TFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWD 600
Qy   966  APNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGC 1025
          ||||||||||||| ::||||||||||||||||||||||||||||||||||||||||
Db   601  APNGTPNPNPNPGT-TLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGC 659
Qy  1026  VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDA 1085
          |||||||||||||||||||||||||:| :|||||||||||||||||||||||||||||
Db   660  VYSESFGLGGIGVLTPDNYLITPALDLLANGGKLTFWVCAQDANYASEHYAVVASSTGNDA 719
Qy  1086  SNFTNALLEEITTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDL 1145
          ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db   720  SNFTNALLEEITTAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDL 779
Qy  1146  DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSN 1205
          ||||||||||||||||||||||||||||||||||||||| |||||||||||||||:|
Db   780  DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGTN 839
Qy  1206  VVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNGFPGDHYAVMISKTGTNAGDFTV 1265
          ||:||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   840  VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNGFPGDHYAVMISKTGTNAGDFTV 899
Qy  1266  VFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   900  VFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 959
Qy  1326  DDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   960  DDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1019
Qy  1386  KKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFE-NGIPASWKT 1444
          :||:::|||:|||||||:|| :  |::||||:||:||.:::||||||: |:||.:
Db  1020  KVCVNVTINPTQFNPVQNLKAQ--PDGGDVVLKWEAPSGKRGELLNEDFEGDAIPTGWTA 1077
Qy  1445  IDADGDGNNW-TTTPPGG----SSFAGHNSAICVSSASHINFEGQNPDNYLVTPEL 1497
          :|||||||||  : :  ::|:::|.||::||||||||||||||||||||||||||
Db  1078  LDADGDGNNWDITLNEFTRGERHVLSPLRASNVAISYSSLLQGQEYLPLTPNNFLITPKV 1137
Qy  1498  SLPGGGTLTFWVCAQD-ANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAI 1556
          ||||||||||||||||  |||:|||||||||:|:|    :::||||||:|||:|||
Db  1138  E--GAKKITYKVGSPGLPQWSHDHYALCISKSGTAADF------EVIFEETMTYTCQGGA 1189
Qy  1557  RGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVIT-SGNAPSYTYTIYRNN 1615
          ::||::|||||||||:||||||||||||:::|||||:||:||: |:||.||||||:
Db  1190  NLTREK------DLPAGTKYVAFRHYNCTDVLGIMIDDVVITGEGEGPSYTYVYRDG 1241
Qy  1616  TQIASGVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTV 1675
```

```
Db  1242  TKIQEGLTETTYRDAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTV  1301
               |:|  :||||||      ::  :|:|||||| ||||||: |:| :   ||  |:||||||
Qy  1676  VGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK     1732
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||:|
Db  1302  VGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGYYAVMVVVDGKSYVEKLAIK     1358

Db     6  LLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETKGGTFA   65
              : .:||||  :|       : : . :    .   .:|    : :   ::   ::  |:
Qy    10  IALCSSLLGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGI--   65

Qy    66  SVSIPGAFPT----GEVGSPEVPAVRKLIAVPVGATPVVRVKS--FTEQVYSLNQYG    116
                      .:   :  : :  .  :  :::.|: .  :   ::   :
Db    66  -----GQVPYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL----  116

Qy   117  SEKLMPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINP  176
                 .:  . ::   ::     | :   .:          : :    |::    |  | :
Db   117  --IAPSKGMIRNEDPKKIPYV-GKTYSQNKFFPGEIATLDDPFILRDVRGQVVNFAP    172

Qy   177  VQYDVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTD  233
          :|| . : |:|: . :||   :  :.|    ||  : .:   .  |  |:   |      :
Db   173  LQYNFVTKIRIYTEITVAVSETSEQGKNLNKKGTFAG------FEDTYKRMF-----MN   222

Qy   234  HGDLYNTPV-----RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASI   287
             : |.|||      ||:|: |.:  |   . |   |:|   | : .:  | :   |.:|
Db   223  YEPGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI  282

Qy   288  KAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYYSAVDGDYF  341
             |: :.||     : | :|:|   :|:::||. :|.   :|:    : |   :.:.||:|
Db   283  QQFVKQEYEKEGND---------LTYVLLIGDHKDIPAKITPGIKSDQV---YGQIVGNDHY 332

Qy   342  PEMYTFRMSASSPELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK   401
           | :  .|.| ::|:|  | .::  || .| ||    ::.  |:|  :   ::     .:|
Db   333  NEVFIGRFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGBSDIQ  392

Qy   402  Y-GMQYYNQEHGYTDVYNYLKAPYTGCY--------SHLNTGVSFANYTAHGSETAW    450
          : :  :||:| |||  :   |  .|: :          .:| |  ||||.|||||||
Db   393  HENVIANLLTQYGYTKIIK---------CYDPGVTPKNIIDAFNGGISLANYTGHGSETAW  444

Qy   451  ADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV----KEKGAYAYIG  506
          . |  .|::| :|||.::  : .:|:| . .:  .:|||:|::     |||   |  :|
Db   445  GTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIA  504

Qy   507  SSPNSYWGEDYHWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH  566
          | .:| |  :      |.                  |::   |           :    :
Db   505  STINQSW--------ASPMRG------------QDEMNEI---------LCEKH        529

Qy   567  AGNI------GNITHIGAHYYWEAYH---------VLGDGSVMPYRAMPKTNTYTLPASL  611
          ..|:       ::|: ||  .: : :         :| |  .|   | : |  | .||  :
Db   530  PNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQI  589

Qy   612  PQNQASYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN  670
           :.:|  . :.| :   :| ..|: . .:. :| :| | ::.  .    .| |:||: ||
Db   590  NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGYN  647

Qy   671  YLPVTKQIQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK---KAEGSREVKRIGDGL  726
          .. : |  :. ||.:|||||||||||||||||||||:||| |   ||:   |  :|| |
Db   648  KETVIKTINTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELV  707

Qy   727  FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA  783
          .:: .:||: || :|  :|| |:|||:: . || :|||||:  |.||.  |  :   :
Db   708  LLSVSDAPELLRSGQAEIVLEAHDVWNDSGSYQIILDADHDQYGQVIPSDTHTLWPNCSV  767

Qy   784  SSNLYSANFEYLVPANADPVVTTQNLIVTGQGEVVIPGGVTVDYCITNPEPASGKMWIAGD  843
          .:||  :||.| ||||||| |:.. |:: ::   .|.|  |: |  ::.|. | | |||.
Db   768  PANLF-APFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQ  825

Qy   844  GGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLT  903
          |.  |  :||.|||||||| | :::.|.|||| :::| ..| .:|||||||||||||||
Db   826  G---PTKEDDYVFEAGKKYHFLMKKMGSGDGTELTISEGGGSDYTYTVYRDGTKIKEGLT  882

Qy   904  ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK  963
          ||||||||||.|||||||||||||||||||||||||||||||||||||||||.|||||||
Db   883  ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK  942

Qy   964  WDAPNGTPNPNPNPNPNPG---TTLSESFENGIPASWKTIDADGDGHWKPGNAPGIAGY  1020
          |||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db   943  WDAPNGTPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHWKPGNAPGIAGY   1002
```

RESULT 8
Q51839

```
ID   Q51839       PRELIMINARY;       PRT;   1706 AA.
AC   Q51839; Q51840;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arginine-specific thiol protease precursor.
GN   PRTR.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=95160709; PubMed=7857299;
RA   Kirszbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N.,
RA   Reynolds E.C.;
RT   "Complete nucleotide sequence of a gene prtR of Porphyromonas
RT   gingivalis W50 encoding a 132 kDa protein that contains an arginine-
RT   specific thiol endopeptidase domain and a haemagglutinin domain.";
RL   Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Reynolds E.;
RX   MEDLINE=96311339; PubMed=8713096;
RA   Slakeski N., Cleal S.M., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtR that encodes
RT   an arginine-specific thiol proteinase and multiple adhesins.";
RL   Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Reynolds E.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Slakeski N.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L26341, AAC18876.1; -.
DR   HSSP; P95493; 1CVR.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease; Signal.
FT   SIGNAL     1     23       POTENTIAL.
FT   CHAIN    228   1706       ARGININE-SPECIFIC THIOL PROTEASE.
SQ   SEQUENCE   1706 AA;  185626 MW;  E8BDF07C9813B844 CRC64;
```

Query Match   46.5%;   Score 4272;   DB 2;   Length 1706;
Best Local Similarity 50.9%;   Pred. No. 1.4e-212;
Matches 932; Conservative 204; Mismatches 456; Indels 240; Gaps 47;

```
RL   J. Biol. Chem. 270:1007-1010(1995).
DR   EMBL; U15282; AAA69539.1; -.
DR   PIR; A55426; A55426.
DR   HSSP; P95499; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   CHAIN       228    719   MATURE 50-KDA CYSTEINE PROTEINASE
FT                            GINGIPAIN.
SQ   SEQUENCE   1704 AA;   185436 MW;   6A34B40131C2A676 CRC64;

Query Match            46.5%;   Score 4270;   DB 2;   Length 1704;
Best Local Similarity  50.9%;   Pred. No. 1.8e-212; -.
Matches   932;   Conservative 204;   Mismatches 456;   Indels 238;   Gaps 47;

Qy      6 LLIAASLLGVGLYAQSAKIKLDAPTRTTCNNSFKQPDASFSFNEVELTKVETKGGTFA   65
          :.:||||||   :.:   : ||    ::.:::| :.:   :::  : ||  ||
Db     10 IALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVIK--VQFRMDNLKFTEVQTPKGI-- 65

Qy     66 SVSIPGAFPT--------GEVGSPEVPAVRKLIAVPVGATPVVRVKS--FTEQVYSLNQYG 116
                         ||||  :      |    :.|:   .:| : |::|  |:.|.::.:|:
Db     66 ------GQVPTYEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL----- 116

Qy    117 SEKLMPHIQPSMSKSDDPEKVPYYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINP 176
          : ::  | : :||::||  ||  || :   | :||.|:||  ::  :||| .: |  :|.|
Db    117 ---IAPSKGMIRNEDDPKKIPYVY-GKSYSQNKFFPGEIATLDDPFILRDVRGQVVNFAP 172

Qy    177 VQYDVVANQLKVRNNIEI--EVSFQGADEVATQRLYDASFSYFETAYKQLFNRDVVTD 233
          :||:|   :|: ...: :  | : ||.:  . :.:   :.  | :  | :|  :   :.:
Db    173 LQYNPVKTLRIYTEITVAVSETSEQGKNILNKKGTFAG-----FEDTYKRMF------MN 222

Qy    234 HGDLYNTPV--------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAVGTTNASI 287
          : |.|:||:        || .:| .:  : :.:  .|::|:.|.:: ||.:: ::|:|| |
Db    223 YEPGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI 282

Qy    288 KAFIHKKY-----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYYSAVDGDYF 341
          :.|: ::|     |  | : || ::|.||||.:|| .:  .|:|  : ||:|::.|:|::|:
Db    283 QQFVKQEYEKEGND-------LTYVLLVGDHKDIPAKITPGIKSDQV---YQGIVGNDHY 332

Qy    342 PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK 401
          .|::.:  .:. .:::. .|:|::.|:||| ||::  :: .:||||| ::| |||:.:::
Db    333 NEVFIGRFSCESKEDLKTQIDRTIHYBRNITTEDKWLGQALCIASARGGPSADNGESDIQ 392

Qy    402 Y-GMQYYYNQEHGYTDVYNYLKAPYTGCY-----------SHLNTGVSFANYTAHGSETAW 450
          : | .: :.|.:.||:||:. | | :  :           |  |:|:|:||||:||||||
Db    393 HENVIANLLTQYGYTKIIK------------CYDPGVTPKNIIDAFNGGISLVNYTGHGSETAW 444

Qy    451 ADPLLTTSQLXALTNKDKYFLAIGNCCITAQPDYVQPCFGEVITRV----KEKGAYAYIG 506
          ..     .:|  |||.:  ||.:.| ::. .|| : || :  :    :  :.||  .:|:
Db    445 GTSHFGTTHVKQLTNSNQLPPIFDVACVNGDFLFSMPCFABALMRAQKDGKPTGTVAIIA 504

Qy    507 SSPNSWGEDYYWSVGANAVPGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH 566
          |: |.|.   :.: .::|. .  ||| .  .   |:  :|   ::.. ::  ||: :.|
Db    505 STINQSW---------ASPMRG-----------QDEMNEI------------LCEKH 529

Qy    567 AGNI----GNITHIGAHYWEAYH------------VLGDGSVMPYRAMPKTNTYTLPASL 611
          .||:    | :|.:|||:|:|::             .:|:| :.   .:|  .:|| .||
Db    530 PNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQI 589

Qy    612 PQNQAGSYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN 670

Qy   1021 NSNGCVYSESFGLGGIGVLITPDNYLITPALDLPNGGKJLTFWVCAQDANYASEHYAVVASS 1080
          ||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db   1003 NSNGCVYSESFGLGGIGVLITPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASS 1062

Qy   1081 TGNDASNFTNALLEEITTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
          |||||||||||||||||||||||||:|:.|||||||||||||||||||||||||||||||
Db   1063 TGNDASNFTNALLEEITTAKGVRSPEAMRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1122

Qy   1141 FYIDLDEVEIKANGKRADFTEFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1123 FYIDLDEVEIKANGKRADFTEFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTA 1182

Qy   1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260
          |||:||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db   1183 HGGTNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA 1242

Qy   1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
          |||||||||||||||||||||||||||.||||||||||||||||||||||||||||||||
Db   1243 GDFTVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1302

Qy   1321 NYILLDDIQFTMGGSPTPDTYTYTVVRDGTKIKEGLTETTFBEDGVATGNHEYCVEVKYT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1303 NYILLDDIQFTMGGSPTPDTYTYTVVRDGTKIKEGLTETTFBEDGVATGNHEYCVEVKYT 1362

Qy   1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE------------ 1428
          ||||||||:|:||||||||||:||.|:  | :.. :|||:||.:|:     .::|:
Db   1363 AGVSPKKCVNVTVNSTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTEGSREVKRIGDGL 1420

Qy   1429 --------------VLNEDFENGIPASWKTIDADGDGNNWTTTPPPGGSSFAGHNSA 1471
                        ||.:::  .: ::|  |:||:|:: .   .  :.::  .   ::
Db   1421 FVTIEPANDVRANEAKVVLAADNVWGDNTGYFL-LDADHNTFGSVIPATGPLFTGTASS 1479

Qy   1472 ICVSSASHINFEG--PQNPD-----NYLVT--PELSLPGGGTLTFWVCAQDANYASEHY 1521
          .: :   : :::|  |:| :     ::|:|  :   : .   :|.|   |:::  :.:
Db   1480 DLYSA----NFESLIPANADPVVTTQNIIVTGQGEVVIPGG----VVDYCITNPEPASGKM 1532

Qy   1522 AVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-A 1580
          ..:|.    :  .: :|  .::| : .:  : .::.  |:  .:| ::   .|  .  :
Db   1533 WI-AGDGGNQPARYDFTFE----------------------------AGKXYTFT 1559

Qy   1581 FRHFGCTDFFWINLDDVVITSGNAPSYTYTYRNNTQIASGVTETYRDDPLATGFYTYG 1640
          .|: |.|: : .:  ::  ::. .||||||.|  .::..|.||| ||:.:.:. :::||
Db   1560 MRRAGMGDG------TDMEVEDDSPASYTYTVVRDGTKIKEGLTETTYRDAGMSAQSHEYC 1614

Qy   1641 VKVVYPNGESAIETATINITSLADVTAQKPYLTVVGKTITVTCQGEAMIYDMNGRRLAA 1700
          |:|.|. ||..  .|:  ::::|||||||||.|:|||||:|| |||||||||||||||||
Db   1615 VEVKYTAGVSPKVCVDYIPDGVADVTAQKPYTILTVVGKTIIVTVCQGEAMIYDMNGRRLAA 1674

Qy   1701 GRNTVVYTAQQGHYAVMVVVDGKSYVEKLAVK 1732
          |||||||||||||:||||||||||||||||:|
Db   1675 GRNTVVYTAQQGYYAVMVVVDGKSYVEKLAIK 1706
```

RESULT 9
Q51816

```
ID   Q51816     PRELIMINARY;     PRT;     1704 AA.
AC   Q51816;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arg-gingipain-1 proteinase.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroides; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95138080; PubMed-7836351;
RA   Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
RA   Travis J., Barr P.J.;
RT   "Molecular cloning and structural characterization of the Arg-
RT   gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a
RT   proteinase-adhesin polyprotein.";
```

```
Db    590  NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGYN 647

Qy    1703 NTVVVTAQGHYAVMVVVDGKSYVEKLAVK 1732
           ||||||||||||| ||||||||||||||||
Db    1675 NTVVVTAQGGYAVMVVVDGKSYVEKLAVK 1704

RESULT 10
Q51838
ID   Q51838        PRELIMINARY;      PRT;   1706 AA.
AC   Q51838;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Protease precursor.
GN   PRPR1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96071894; PubMed=7591131;
RA   Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.;
RT   "Characterization, genetic analysis, and expression of a protease
RT   antigen (PrpRI) of Porphyromonas gingivalis W50.";
RL   Infect. Immun. 63:4744-4754(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.;
RT   "The prpR1 and the prR2 arginine-specific protease genes of
RT   Porphyromonas gingivalis W50 produce five biochemically distinct
RT   enzymes.";
RL   Mol. Microbiol. 23:0-0(1997).
DR   EMBL; X82680; CAA57997.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Signal.
FT   SIGNAL          1      23       POTENTIAL.
FT   CHAIN         228     719       ALPHA-PROTEASE.
FT   CHAIN         720    1262       BETA-ADHESIN.
SQ   SEQUENCE   1706 AA;  185705 MW;  0E56DCD87FDA8CDD CRC64;

Query Match             46.4%;  Score 4256;  DB 2;  Length 1706;
Best Local Similarity   50.8%;  Pred. No. 9.7e-212;
Matches  928;  Conservative  206;  Mismatches  465;  Indels  226;  Gaps  45;

Qy       6 LLIAASLLGVGLYAQSAKIKLDAPTRTTCNNSFKQFDASFSFNEVELTKVETKGGTFA 65
           :|| :||||||||||| ||||||| |   |  |: ||:|  :|  |: ||| : ||||
Db      10 IALCSSLLGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGMAQ 67

Qy      66 SVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKS--FTEQVYSLNQYGSEKLMPH 123
           ::||   |||:||  : ::      :   ||    ::||   |:   ::|:|:  :::|:
Db      68 VPTYTEGVNLSEKGMPTLPIISRSLAVSDTREMKVEVVSSKFIEKKNVL------IAPS 120

Qy     124 QPSMSKSDDPEKVPFYVNAAAYARKGFVGDELTQVEMLGTMRGVRIAALTINPVQYDVVA 183
           ::: :|:  :::||:   :    :|  ::|: :|: ::|: ||: ||: :|::| |: |:
Db     121 KGMIMNEDPKKIPYVY-GKSYSQNKFFPGEIATLDDPFILRDVRGQVVNFAPLQYNPVT 179

Db     647  NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGYN 647
Qy     671  YLPVIKQIQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK---KAEGSREVKRIGDGL 726
Db     648  KETVIKTINTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELV 707
Qy     727  FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA 783
Db     708  LLSVSDAPELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQYGGVIPSDTHTLWPNCSV 767
Qy     784  SSNLYSANFEYLVPANADPVVTQNLIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGD 843
Db     768  PANLF-APFEYTVPENADPSCSPTNMIDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQ 825
Qy     844  GGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVRDGTKIKEGLT 903
Db     826  G---PTKEDDYVFEAGKKYHFLMKKMGSGDGTELTISEGGSDYTYTVRDGTKIKEGLT 882
Qy     904  ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK 963
Db     883  ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK 942
Qy     964  WDAPNGTPNPNPNPNPG-TTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNS 1022
Db     943  WDAPNGTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNS 1002
Qy    1023  NGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG 1082
Db    1003  NGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG 1062
Qy    1083  NDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY 1142
Db    1063  NDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY 1122
Qy    1143  IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHG 1202
Db    1123  IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHG 1182
Qy    1203  GSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGD 1262
Db    1183  GTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGD 1242
Qy    1263  FTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYHYNCSDLNY 1322
Db    1243  FTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYHYNCSDLNY 1302
Qy    1323  ILLDDIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG 1382
Db    1303  ILLDDIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG 1362
Qy    1383  VSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE---------- 1428
Db    1363  VSPKECVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTEGSREVKRIGDGLFV 1420
Qy    1429  ----------VLNEDFENGIPASWKTIDADGDGNNWTTTPPPGGSSFAGHNSAIC 1473
Db    1421  TIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSNL 1479
Qy    1474  VSSASHINFEG--PQNPD-----NYLVT--PELSLPGGGTLTFWVCAQDANYASEHYAV 1523
Db    1480  YSA----NFEYLIPANADPVVTQNLIVTGQGEVVIPGG--VYDYCIITNPEPASGKMWI 1532
Qy    1524  YASSTGNDASNFANALLEEVLTAKIVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-AFR 1582
Db    1533  -AGDGGNQPARYDDFTFE-------------------AGKKYTFTMR 1559
Qy    1583  HFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVK 1642
Db    1560  RAGMGDG-----TDMEVEDDSPASYTYTVRDGTKIKEGLTETTYRDAGMSAQSHEYCVE 1614
Qy    1643  VVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGR 1702
Db    1615  VKYAAGVSPKVCVDYIPDGVADVTAQKPYTITVVGKTITVTCQGEAMIYDMNGRRLAAGR 1674
```

```
Qy    184  NQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT  240
           :|::|.||      .::|.| . .|:|:  ||  ||:  .::: ::
Db    180  KTLRIYTEITVAVSETSEQGKNILNKKGTFAG----FEDTYKRMF-----MNYEPGRYT  229

Qy    241  PV-----RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKK  294
           |.       :||| |  |:|  |:|:|.|| || |  ||:.:|||:|::|:||| | |:
Db    230  PVEEKQNGRMTVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAIQQFVKQE  289

Qy    295  Y----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYYSAVDGDYFPEMVTFR  348
           :    |::::|  .|||   .  ||::|.: |.|:|.| .:::|.||.::|||
Db    290  YEKEGND------LTYVLLIGDHKDIPAKITPGIKSDQV---YGQIVGNDHYNEVFIGR  339

Qy    349  MSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKY-GMQYY  407
           ::|:   :::. .||.  |::|| |||:| | :::.||||:.||:||||| .:| || ||
Db    340  FSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQHENVIAN  399

Qy    408  YNQEHGYTDVYNYLKAPYTGCY---------SHLNTGVSFANYTAHGSETAWADPLLTT  457
           : |:.|||:.||. |.|||              |:||||:|||:|||||:|.:|:|
Db    400  LLTQYGYGTKIIK--------CYDPGVTPKNIIDAFNGGISLANYTHGSETAWGTSHFGT  451

Qy    458  SQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV----KEKGAYAYIGSSPNSYW  513
           :  | |:|. ::|  ||:|:|  .  :| .|::| :::.    |:||.|:||.||.||
Db    452  THVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIASTINQSW  511

Qy    514  GEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNI---  570
           :. :   .| .||.|   :||.:  |::||  ||   | | | ::|.||:  :||.|
Db    512  --------ASPMRG------------QDEMNEI------------LCEKHPNNIKRT  536

Qy    571  -GNITHIGAHYWEAYH---------VLGDGSVMPYRAMPKTNIYTLPASLPQNQASY  618
            ||:|.||:|:||||          .||:||.||  ..|||:|.|||:||:|:||:|
Db    537  FGGVTMNGMFAMVEKYKKDGERMLDTWTVFGDPSLLVRTLVPTKMQVTAPQINLTDASV  596

Qy    619  SIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIKQ  677
           ::.|| :|||.| ||||.:||. |||. |||:|||   ||.|||| :|||.|:||:|:|:
Db    597  NVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLILTVVGNKETVIKT  654

Qy    678  IQV-GEPSPYQPVSNLTATTQQQKVTLKWEAPSAK--KAEGSREVKRIGDGLFVTIEPA  733
           |:| ||||||||||||||||||||||.|||.|||   |||||.||.|||||||||||||
Db    655  INTNGEPNPYQPVSNLTATTQQGKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDA  714

Qy    734  ND-VRANEAKVVLAADNVWGDNTGYQVIPSDTHILWPNCSVPANLF-A  790
           |: |||||||||||.|||||.||||::||  | : || ||.|:.|.|
Db    715  PELLRSGQAEIVLEAHDVWNDSGYQILLDADHDQYGQVIPSDTHILWPNCSVPANLF-A  773

Qy    791  NFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCIITNPEPASGKMWIAGDGNQPAR  850
           ||||||||||||:.||.|||||| ||||||..| .: :| ||||:|||: |  |...|:
Db    774  PFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQG---PTK  829

Qy    851  YDDFTFEAGKKYTIFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTEATTFEED  910
           :|||.|||||:|.|:|||||||.||::||.|:||:|||||||||||:||||||| |||||:
Db    830  EDDYVFEAGKKYHFLMKKMKGSGDGTELTISEGGSGDYTYTVYRDGTKIKEGLTETTTFEED  889

Qy    911  GVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGT  970
           ||||||||||||||||||||||||||||||||||||||||||.|||||:|||||||||||
Db    890  GVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGT  949

Qy    971  PNPNPNPNPNPG---TTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCVY  1027
           |||||||||||    |::|||||||||||||||||||||||||||||||||||||||||
Db    950  PNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCVY  1009

Qy    1028 SESFGLGGIGVLTPDNLITPALDLPNGGKLTFWCAQDANYASEHYAVYASSTGNDASN  1087
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1010 SESFGLGGIGVLTPDNLITPALDLPNGGKLTFWCAQDANYASEHYAVYASSTGNDASN  1069

Qy    1088 FTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDE  1147
           |||||||||||||||||..||||||||||||:|||||||||||||||.|||||||||||
Db    1070 FTNALLEETITAKGVRSPEAIRGRIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDE  1129

Qy    1148 VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVV  1207
           |||||||||||||||||||||||||.|||||||||||||||||||||||||||||||||
Db    1130 VEIKANGKRADFTETFESSTHGEATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVV  1189
```

```
Qy    1208 SSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVF  1267
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1190 SSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVF  1249

Qy    1268 EETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDD  1327
           ||||||||||||||||||||.||||||||||||||||||||||||||||||||||||||
Db    1250 EETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDD  1309

Qy    1328 IQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEBDGVATGNHEYCVEVKYTAGVSPKK  1387
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1310 IQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKK  1369

Qy    1388 CVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE------------  1428
           |||||||||||.| ||||. :    | |:|||:||.|| | |
Db    1370 CVNVTVNSTQFNPVKNLKAQ-PDGGDVLKWEAPSAKKTEGSREVKRIGDGLFVTIEPA  1427

Qy    1429 ----------VLNEDFENGIPASWKTIDADGDGNNWTTPPPGGSSPAGHNSAICVSSAS  1478
                     || |||||||||||||||||||.|.:| |.|||:||:||||
Db    1428 NDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSDLYSA-  1484

Qy    1479 HINFEG--PQNFD---------NYLVT--PELSLPGGGTLTFWVCAQDANYASEHYAVYASST  1528
           |::|     ::|           ||:    .  : |.||.|:| :. || .: .|.. :.|:
Db    1485 --NFEYLIPANADPVVTTQNIIVTGQGEVVIPGG---VYDYCIITNPEPASGKMWI--AGDG  1538

Qy    1529 GNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-AFRHFGCT  1587
           |:.|. .  .: ::.  ||:|::    :    |:||   ||  ||:|||:|  |
Db    1539 GNQPARYDDFTFB-----------------------------AGKKYTFMRRAGMG  1566

Qy    1588 DFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPN  1647
           ||       . .|:::  |||||||.. :..|..|:|||||.  .|.:  |     ::|
Db    1567 DG-----TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVEVKYTA  1621

Qy    1648 GESAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMYDMNGRRLAAGRNTVVY  1707
           |.|.|:.    ||:|||||||||||||||||||||||||||||||||||||||||||||
Db    1622 GVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQGEAMYDMNGRRLAAGRNTVVY  1681

Qy    1708 TAQGGHYAVMVVVDGKSYVEKLAVK  1732
           |||.|:||||||||||||||:|.:|
Db    1682 TAQGGYYAVMVVVDGKSYVKKLAIK  1706
```

```
RESULT 11
ID    P72196
AC    P72196;
DT    01-FEB-1997 (TrEMBLrel. 02, Created)
DT    01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE    TonB-linked adhesin precursor.
GN    TLA.
OS    Porphyromonas gingivalis (Bacteroides gingivalis).
OC    Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC    Porphyromonadaceae; Porphyromonas.
OX    NCBI_TaxID=837;
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=97386416; PubMed=9244265;
RA    Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA    Curtis M.A.;
RT    "The tla gene of Porphyromonas gingivalis W50: a homologue of the
RT    arginine-specific protease precursor (PrpRI) which shares sequence
RT    similarity to TonB-linked receptors.";
RL    J. Bacteriol. 179:4778-4788(1997).
CC    -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR    EMBL; Y07618; CAA68897.1; -.
DR    GO; GO:0019867; C:outer membrane; IEA.
DR    GO; GO:0005524; F:ATP binding; IEA.
DR    GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR    GO; GO:0004872; F:receptor activity; IEA.
DR    GO; GO:0005215; F:transporter activity; IEA.
DR    GO; GO:0006310; P:DNA recombination; IEA.
DR    GO; GO:0006281; P:DNA repair; IEA.
DR    GO; GO:0006260; P:DNA replication; IEA.
```

Qy 1643 VVYPNGBSAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGR 1702
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1308 VVYPNGBSAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGR 1067

Qy 1703 NTVVVTAQGGHYAVMVVVDGKSYVEKLAVK 1732
         ||||||||||||||||||||||||||||||
Db 1068 NTVVVTAQGGHYAVMVVVDGKSYVEKLAVK 1097

RESULT 12
Q9R9B7
ID   Q9R9B7   PRELIMINARY;   PRT;   1687 AA.
AC   Q9R9B7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin/protease.
GN   HAGE.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Dong H., Progulske-Fox A.;
RT   "Cloning and characterization of hagE from P. gingivalis 381.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF026946; AAD01810.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE   1687 AA; 183702 MW; D085B516A399FE70 CRC64;

Query Match   46.2%; Score 4245; DB 2; Length 1687;
Best Local Similarity   51.6%; Pred. No. 3.5e-211;
Matches 920; Conservative 197; Mismatches 443; Indels 224; Gaps 44;

Qy   47 FSFNEVELTKVETKGGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKS-- 104
                   :   : : || ||||: | : |:| :| | :||
Db   30 FRMDNLKFTEVQTPKGMAQVPTYTEGVNLSEKGMPTLPILSRSLAVSDTRMKVEVVSSK 89

Qy   105 FTEQYSLNQYGSEKLMPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTM 164
             | :   ::: ::   : |||:|: | | : : |||
Db   90 FIEKNVL-----IAPSKGMIMRNEDPKKIPYVY-GKSYSQNKFPPGEIATLDDPFIL 141

Qy   165 RGVRIAALTINPVQYDVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETA 221
               :   :| :|:|: ||::| : |
Db   142 RDVRGQVVNFAPLQYNPVTKTLRIYTEITVAVSETSEQGKNILNKKGTFAG-----FEDT 196

Qy   222 YKQLFNRDVVTDHGDLYNTPV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYT 275
            : | :: | | :   :| | : : ||| : | | : |:
Db   197 YKRMF-----MNYEPGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVA 251

Qy   276 DEAEVGTTNASIKAFIHKK-----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVT 329
             : :: | :|| :|||::   :   : | ::|: |: | | |
Db   252 EDIASPVTANAIQQFVKQEYEKEGND------LTYVLLVGDHKDIPAKITPGIKSDQV- 303

Qy   330 DLYYSAVDGDYFPEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLBKVLLIAGADY 389

```
Db  1332  GNHEYCVEVKYTAGVSPKECVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTE  1389
Qy  1429  -----------------------------VLNEDFENGIPASWKTIDADGDGNNWTTTPP  1459
Db  1390  GSREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIP  1448
Qy  1460  PGGSSFAGHNSAICVSSASHINPEG--PQNPD------NYLVT--PELSLPGGGTLTFWV  1509
Db  1449  ATGPLFTGTASSNLYSA---NFEYLIPANADPVVTQNIIVTGQGEVVIPGG---VDY    1501
Qy  1510  CAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKT  1569
Db  1502  CITNPEPASGKWWI-AGDGGNQPARYDDFTFE-----------------------     1532
Qy  1570  VQLPAGTKYV-AFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYR  1628
Db  1533  ----AGKKYTFMRRAGRMGDG-------TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYR  1583
Qy  1629  DPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKEYTLTVGKTITVTCCQEA  1688
Db  1584  DAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKIITVTCCQEA  1643
Qy  1689  MIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK  1732
Db  1644  MIYDMNGRRLAAGRNTVVYTAQGGYYAVMVVVDGKSYVEKLAVK  1687
```

RESULT 13

```
P96966
ID   P96966    PRELIMINARY;    PRT;    364 AA.
AC   P96966;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cysteine protease.
GN   KGP-381.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Lepine G., Whitlock J., Wojciechowski L., Progulske-Fox A.;
RT   "Cloning, sequencing and characterization of hagD, a member of the
RT   HArep mnltigene family in Porphyromonas gingivalis.";
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U68468; AAB49690.1; -.
DR   MEROPS; C25.002; -.
GO   GO:0008234; F:cysteine-type peptidase activity; IEA.
GO   GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR001769; Peptidase_C25.
DR   Pfam; PF01364; Peptidase_C25; 1.
KW   Protease.
SQ   SEQUENCE   364 AA;   40161 MW;   33DB6AB1CF57557B CRC64;

Query Match             364 AA;   Score 1854;   DB 2;   Length 364;
Best Local Similarity   20.2%;   Pred. No. 3e-88;
Matches  363;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy     1  MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK   60

Qy    61  GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYVSLNQYGSEKL  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYVSLNQYGSEKL  120

Qy   121  MPHQPSMSKSDDPEKVPFVVNAAAYARKGPVGQELTQVEMLGTMRGVRIAALTINPVQYD  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  MPHQPSMSKSDDPEKVPFVVNAAAYARKGPVGQELTQVEMLGTMRGVRIAALTINPVQYD  180

Qy   181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT  240
```

```
Db   304  --YGQIVGNDHYNEVFLGRFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEG  361
Qy   390  SWNSQVGQPTIKY-GMQYYVNQEHGYTDVYNYLKAPYTGCY-----------SHLNTGVSF  438
Db   362  GPSADNGESDICHENVIANLLTQYGYTKIIK--------CVDPGVTPKNIIDAFNGGISL  413
Qy   439  ANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDVVQPCFGEVITRV--  496
Db   414  VNYTGHGSETAWGTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQK  473
Qy   497  --KEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVN  554
Db   474  DGKPTGVAIIASTINQSW------------------ASPMRG------------QDEMNEI  505
Qy   555  SIMWAGNLAATHAGNI-----GNITHIGAHYIWEAYH-----------VLGDGSVMPYRAM  599
Db   506  --------LCEKHPNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLV  558
Qy   600  PKTNTVTLPASLPQNQASYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMKQITE  658
Db   559  PTKMQVTAPAQINLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTN  616
Qy   659  NGNYDVVITRSNYLPVIKQIOV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK---KAE  714
Db   617  ESTLITLVGYNKETVTKTINTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTN  676
Qy   715  GSREVKRIGDGLFVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIP  773
Db   677  TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWNDSGYQLLDADHDQYGQVIP  736
Qy   774  A-TGPLFTG-TASSNLYSANFEYLVPANADPVVTQNIIVTGQGEVVIPGGVYDYCITNP  831
Db   737  SDTHTLWPNCSVPANLF-APFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAP  795
Qy   832  EPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTV  891
Db   796  Q-ANAKIWIAGQG---PTKEDDYVFEAGKKYHFLMKKMGSGDGTELITISEGGSDYTYTV  851
Qy   892  YRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNL  951
Db   852  YRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNL  911
Qy   952  TGSSVGQKVTLKWDAPNGTENPNPNPNPG---TTLSESFENGIPASWKTIDADGDGHG  1008
Db   912  TGSAVGQKVTLKWDAPNGTENPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHG  971
Qy  1009  WKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDAN  1068
Db   972  WKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDAN  1031
Qy  1069  YASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTK  1128
Db  1032  YASEHYAVYASSTGNDASNFNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTK  1091
Qy  1129  YVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL  1188
Db  1092  YVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL  1151
Qy  1189  CLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDH  1248
Db  1152  CLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDH  1211
Qy  1249  YAVMLSKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERRTVDLPAGTK  1308
Db  1212  YAVMLSKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIRRTVDLPAGTK  1271
Qy  1309  YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVAT  1368
Db  1272  YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVAT  1331
Qy  1369  GNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAIILKWNAPASKRAE  1428
```

```
Db    181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    296  SSTVCGTLHYATDAILYENFENGPVPNGWLVIDADGDGFSW-----GHYLNAYDAFPGHN  350

Qy    241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1203  GSNVVSSFSW-NGMA-LNPDNYLISKDVTGATKVKYYAVND-GFPGDHYAVMLSKTGTN  1259

Db    241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351  GGHCSLSASYVPGIGPVTPDNYLITPKVEGAKRVKYWVSTQDANWAAEHYAVMASTTGTA  410

Qy    301  ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1260  AGDPTVVPEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSD  1319

Db    301  ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411  VGDFVILFEETMT----------AKPTGAWYERTINLPEGTKYIAWRHYNCTD  453

Qy    361  IDK  363
           |||
Qy   1320  LNYILLDDIQF--TMGGSPTP-TDYT  1342
                       :.: |||||::|
Db    361  IDK  363
           |||
Db    454  IYFLKLDDITVFGTPASEPEPVTDFVVSLIENNKGRLKWNYPNGYEPDKTDDKDPLQLAG  513

Qy   1343  YTVVRDGT----------KIKEGLIET-TFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTV  1393
                  :.:          :.||::::.  :.::
Db    514  YNIYANGSLLVHIQDPTVLEYIDETYSSRDDQVEV--EYCVTAVYNDNIESQSVCDKLI  570

Qy   1394  NSTQFNPVQNLTAEQAPNSMDALLKWNAPASKRAEVLNEDFENG-IPASWKTIDADGDGN  1452
           :::|||:|::.:: :|||::|   |::|  |:   :|:|   ||| |||||||||||||||
Db    571  YDSQSD----------------------IILYEGFEAGSIPEGWLLIDADGDNV  602

Qy   1453  NWTTTPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLPGGGTLTPWVCAQ  1512
           ||      :::.: ::: :::|     |||   |:||||:||| ||| |  :|.|||||
Db    603  NWDYYP--WTMYGHDSBKCIASPSYLPMIGVLTPDNYLVTPLE--GAKLVKYWVSAQ  656

Qy   1513  DANVASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQL  1572
                :||:||||||||||   || :|||  ::|||:      :||||||:|||:||.|
Db    657  DAVYSAEHYAVMVSTTGTAVEDFV-LLFEETMTAK---------ANGAWYERTITL  702

Qy   1573  PAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYIYRNNTQIASGVTE---------  1624
           |||||||||||||||||         :::|  |||||||||||||:|:.::
Db    703  PAGTKYIAWRHYDCTDMFILLLDDI------------TVRSTETVPEPVTDFVVSLIEN  750

Qy   1625  -------------------------TTYRDDPLATGFYTGVKVVYPNGESAIETATLNITSLADVT-  1666
                                     ||:|||||||||||||||||||::||.|:||::::|:||||
Db    751  NKGRLKWNYPNGYEPDKTDDKKPIQLTGYN--------IYANGSLLVHIQDPTVLEYIDETY  804

Qy   1667  -----AQKPYTLTVV-----------------------GKTITV--TCQGEAMI-----  1696
                ||:|||:|||                       ||:||.  :||||:|  YDMNGR
Db    805  SSRDGQVEMEYCVTAVYNDNIESQSVCDKLNYTITSLDNIQSDTSLKIYPNPASYVVRIE  864

Qy   1697  RLAAGRNTV-VYTAQG-----------------------GHYAVMVVVDGKSYVEKLAVK  1732
                                                   |:||.|:|.||:.||::.|:
Db    865  GLSRSKSTIELYNALGICILREETHSEKTEIDVSRLNDGVYLIKVVGGNKTTTRKVEIK  923
```

RESULT 15
Q9KIB3
ID   Q9KIB3   PRELIMINARY;   PRT;   312 AA.
AC   Q9KIB3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical outer membrane protein PG27.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C.;
RT   "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
```
Qy4J0
ID    Q9F4J0      PRELIMINARY;      PRT;      925 AA.
AC    Q9F4J0;
DT    01-MAR-2001 (TrEMBLrel. 16, Created)
DT    01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT    01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE    Putative outer membrane protein PG57.
OS    Porphyromonas gingivalis (Bacteroides gingivalis).
OC    Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC    Porphyromonadaceae; Porphyromonas.
OX    NCBI_TaxID=837;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=W50;
RA    Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA    Hocking D., Webb E.;
RT    "P. gingivalis polypeptides and nucleic acids.";
RL    Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=W50;
RA    Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA    Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT    "Identification of vaccine candidates from genomic analysis of
RT    Porphyromonas gingivalis.";
RL    Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AY007522; AAG24228.1; -.
DR    InterPro; IPR003961; FN_III.
DR    SMART; SM00060; FN3; 1.
SQ    SEQUENCE   925 AA;   103632 MW;   5FF2198D6514DAE8 CRC64;
```

```
Query Match             10.7%;   Score 980.5;   DB 2;   Length 925;
Best Local Similarity   28.4%;   Pred. No. 2 4e-42;
Matches  289;   Conservative 109;   Mismatches 264;   Indels 357;   Gaps   42;

Qy    960  VTLKWDA-PNGTPNPNPNPN------------PGTTLSESFENG-IPASWKTIDADGDGHGWK  1010
           |.||:.: |||:|::|:|:|            :| :||||:.::: |||||||||||||||
Db     16  ILLSWAAITNPTAQEISGMNASCLAAPAQPDTILYESFENGPVPNGWLEIDADGDATWG  75

Qy   1011  PGNAPGLAGYNSNG-CVYSESFGLGGIGVLTPDNYLITPALDLPNGKLTFWVCAQDANY  1069
           |||:|| ||.:: :  :.:|||.| |:|:|||||||:||    ::: |:.|:.|:
Db     76  SPSGSFSVPYGHNGLCTYSHIRS---GISTAGNYLITP--NIEGAKRVKYWVCNQ---Y  126

Qy   1070  AS--EHYAVYASSTGNDASNFTNALLEETIITAKGVRSPKAIRGRIQGTWRQKTVDLPAGT  1127
           |:  |:||||||||||| :||  ::|::: ||| :|   |:: .|:|||| |.|:.:|||
Db    127  STNPEHYAVMVSTTGTAIEDFV-LLFDSITGK----PTPL-----VWRRRIVDLPEGT  175

Qy   1128  KYVAFRHFQSTDMF--YIDLDEV----------------------------------  1148
           |||||||.|:|:: :  ::||:::
Db    176  KYIAWRHYKVTDSHTEFLKLDDVTVYRSIEGPEPATDFTVINIGQNVGRLTWNYPEDYQP  235

Qy   1149  -------------EIKANGK-------------------------------RADFTETFE  1164
                        :|.:||:                               |.|::|:|
Db    236  EGKGNEELQLSGYNIYANGTLLAQIKDVSILEVDSTYSLRDNPLQVEYCVTAVYDESIE  295

Qy   1165  SST---------------------HGEAPAEWTTIDADGDGQGWLCLSSGQ---LDWLTAHG  1202
```

DR   EMBL; AF237555; AAF81413.1; -.
SQ   SEQUENCE   312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;

Query Match
Best Local Similarity   7.9%;   Score 722; DB 2; Length 312;
Matches 150; Conservative 49; Mismatches 97; Indels 20; Gaps 7;

Qy   1430 LNEDFENGIPASWKTIDADGDGNNWTTPPGGSSFAGHNSAICVSSASHINFEGPQNPD 1489
                 |:|||||||||||||||| ::|:||||||:|
Db   4 LSESFBSGIPAIWKTIDADGDGYNWMHL----TNFTG-QSGLCVSSASYIGGVGALTPD 57

Qy   1490 NYLVTPELSLPGGG--TLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEVLTAK 1547
                 |||:||| || :||| ||||:||||||||:| || ||||:|||||
Db   58 NYLITPELKLPTDALVEIIYWVCTQDLTAPSEHYAVYSSSTGNNAADFVNLLYEETLTAK 117

Qy   1548 TVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAP-- 1605
                 :|||| |||||:|||:::|| :|||||| |:|||:|:
Db   118 R-IQSPELIRGNRTQGVWYQRKVVLPNDTKYVAFRHFNSTDNFWLNLDEVSILYTPLPRR 176

Qy   1606 -------SYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLN 1658
                        |||::||:|||||:   :|:|| :|:|
Db   177 APCPHPGGYTYSVFRDGQKIASGLSALAYIDTDVPYGTQDYCVQVNYLQGDSYKVCKNIV 236

Qy   1659 ITSLADV-TAQKPYTLTVVGKTIITVTC-QGEAMIYDMNGRRLAAGRNTVVYTAQCGHYAV 1716
                 ||:|| |:|||||:||||||| :||:|| :|:|| ||:||:||:|| |:
Db   237 VANSANIYGADKPFALTVVGKTIVASAFKGEITLYDIRGRLIASGCDTLRYKAENGFYLI 296

Qy   1717 MVVVDGKSYVEKLAVK 1732
                 :|:| |||:: ||::
Db   297 KIQVNGTVVTEKIQIQ 312

Search completed: May 18, 2004, 11:35:11
Job time : 75.1557 secs

OM protein - protein search, using sw model

```
                Copyright  GenCore version 5.1.6
                Copyright (c) 1993 - 2004 Compugen Ltd.
```

Run on: May 18, 2004, 11:30:39 ; Search time 28.7501 Seconds
                                      (without alignments)
                                      3110.116 Million cell updates/sec

Title: US-08-353-485-10
Perfect score: 9179
Sequence: 1 MRKLLLIAASLLGVGLYAQ.........HYAVMVVVDGKSYVEKLAVK 1732

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: ,   389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued Patents AA:*
        1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
        2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
        3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
        4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
        5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
        6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result   %
No.   Score Match Length DB   ID   Description
----- ----- ----- ------ ---- --   -----------
1   9179 100.0 1732   2   US-08-570-311-10   Sequence 10, Appl
2   9179 100.0 1732   2   US-08-353-485-10   Sequence 10, Appl
3   9167 99.9 1732   4   US-09-066-330-11   Sequence 11, Appl
4   6151.5 67.0 1358   2   US-08-570-311-27   Sequence 27, Appl
5   4811 52.4 942   1   US-08-141-324-14   Sequence 14, Appl
6   4811 52.4 942   1   US-08-541-902-14   Sequence 14, Appl
7   4673.5 50.9 1087   2   US-08-570-311-8   Sequence 8, Appli
8   4673.5 50.9 1087   2   US-08-353-485-8   Sequence 8, Appli
9   4272 46.5 1706   4   US-09-066-330-10   Sequence 10, Appl
10   4270 46.5 1704   3   US-08-336-308A-10   Sequence 10, Appl
11   4270 46.5 1704   3   US-08-822-324-6   Sequence 6, Appli
12   4270 46.5 1704   3   US-09-490-931-10   Sequence 10, Appl
13   4245 46.2 1687   2   US-08-570-311-29   Sequence 29, Appl
14   4217 45.9 2628   2   US-08-570-311-14   Sequence 14, Appl
15   4135.5 45.1 1477   4   US-09-482-500A-1   Sequence 1, Appli
16   2565 27.9 509   3   US-08-822-324-8   Sequence 8, Appli
17   2034.5 22.2 456   2   US-08-570-311-18   Sequence 18, Appl
18   2034.5 22.2 456   2   US-08-570-311-20   Sequence 20, Appl
19   2025.5 22.1 439   2   US-08-570-311-22   Sequence 22, Appl
20   2011.5 21.9 450   2   US-08-570-311-16   Sequence 16, Appl
21   1854 20.2 364   2   US-08-570-311-26   Sequence 26, Appl
22   812.5 8.9   497   3   US-08-570-311-2   Sequence 2, Appli
23   812.5 8.9   497   2   US-08-353-485-2   Sequence 2, Appli
24   698   7.6   737   1   US-08-119-361-5   Sequence 5, Appli
25   698   7.6   737   3   US-08-336-308A-4   Sequence 4, Appli
26   698   7.6   737   3   US-08-822-324-4   Sequence 4, Appli
27   698   7.6   737   3   US-09-490-931-4   Sequence 4, Appli
28   526   5.7   US-09-482-500A-2   Sequence 2, Appli
29   492   4.5   US-09-482-500A-3   Sequence 3, Appli
30   412.5 3.4   US-09-134-000C-5999 Sequence 5999, Ap
31   309   3.3   US-09-328-352-5503   Sequence 5503, Ap
32   306   3.3   US-09-252-991A-30227 Sequence 30227, A
33   304.5 3.2   US-09-071-035-258   Sequence 258, App
34   293   3.2   US-09-071-035-262   Sequence 262, App
35   293   3.0   US-09-071-035-266   Sequence 266, App
36   293   2.7   US-09-543-681A-5434 Sequence 5434, Ap
37   277   2.7   US-09-268-347-36   Sequence 36, Appl
38   250.5 2.7   US-09-328-352-5486   Sequence 5486, Ap
39   246.5 2.6   US-09-377-155-33   Sequence 33, Appl
40   237.5 2.6   US-08-913-942-4   Sequence 4, Appli
41   237.5 2.6   US-09-669-974-33   Sequence 33, Appl
42   237.5 2.6   US-09-797-862-33   Sequence 33, Appl
43   236   2.6   US-08-822-324-18   Sequence 18, Appl
44   236   2.6   US-08-107-532A-4706 Sequence 4706, Ap
45   235.5 2.6   US-08-409-995-4   Sequence 4, Appli
```

ALIGNMENTS

RESULT 1
US-08-570-311-10
; Sequence 10, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;; APPLICANT:  Progulske-Fox, Ann
;; APPLICANT:  Tumwasorn, Somying
;; APPLICANT:  Lepine, Guylaine
;; APPLICANT:  Han, Naiming
;; APPLICANT:  Lantz, Marilyn
;; APPLICANT:  Patti, Joseph
;; TITLE OF INVENTION:  Cloned Porphyromonas gingivalis Genes
;; TITLE OF INVENTION:  and Probes for the Detection of Periodontal Disease
;; NUMBER OF SEQUENCES:  29
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE:  Ted W. Whitlock
;; STREET:  2421 N.W. 41st Street, Suite A-1
;; CITY:  Gainesville
;; STATE:  FL
;; COUNTRY:  USA
;; ZIP:  32606
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE:  Floppy disk
;; COMPUTER:  IBM PC compatible
;; OPERATING SYSTEM:  PC-DOS/MS-DOS
;; SOFTWARE:  PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER:  US/08/570,311
;; FILING DATE:
;; CLASSIFICATION:  424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:  US 08/353,485
;; FILING DATE:  09-DEC-1994
;; CLASSIFICATION:  424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:  US 07/647,119
;; FILING DATE:  25-JAN-1991
;; CLASSIFICATION:  424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:  US 07/241,640
;; FILING DATE:  08-SEP-1988
;; ATTORNEY/AGENT INFORMATION:
;; NAME:  Whitlock, Ted W.
;; REGISTRATION NUMBER:  36,965
;; REFERENCE/DOCKET NUMBER:  UF15.C3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE:  (904) 375-8100
;; TELEFAX:  (904) 372-5800
;; INFORMATION FOR SEQ ID NO:  10:
;; SEQUENCE CHARACTERISTICS:

```
;
;   LENGTH: 1732 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-10

Query Match       100.0%;      Score 9179;    DB 2;    Length 1732;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 1732;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MRKLLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRKLLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK 60

Qy     61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRKSFTEQVYSLNQYGGEKL 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRKSFTEQVYSLNQYGGEEKL 120

Qy    121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180

Qy    181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240

Qy    241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300

Qy    301 ASAAPVFLALVGDTDVLSGEKGKKTKKVTDLYYSAVDGDYPEMYTFRMSASSPEELTNI 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ASAAPVFLALVGDTDVLSGEKGKKTKKVTDLYYSAVDGDYPEMYTFRMSASSPEELTNI 360

Qy    361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNY 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNY 420

Qy    421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTSQLKALTNKDKYFLAIGNCCITA 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTSQLKALTNKDKYFLAIGNCCITA 480

Qy    481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS 540
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS 540

Qy    541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP 600
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP 600

Qy    601 KINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 KINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660

Qy    661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720

Qy    721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Qy    781 GTASSNLYSANFEYLVPANADPVVTQNIIVTGQGEVVIPGGVYDYCIITNPEPASGKMWI 840
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTASSNLYSANFEYLVPANADPVVTQNIIVTGQGEVVIPGGVYDYCIITNPEPASGKMWI 840

Qy    841 AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900

Qy    901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960

Qy    961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020

Qy   1021 NSNGCVYSESFGLGGIGVLITPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 NSNGCVYSESFGLGGIGVLITPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080

Qy   1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140

Qy   1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDQGWLCLSSGQLDWLTA 1200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDQGWLCLSSGQLDWLTA 1200

Qy   1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNA 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNA 1260

Qy   1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320

Qy   1321 NVILLDDIQFTMGGSPTPTDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 NYILLDDIQFTMGGSPTPTDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380

Qy   1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440

Qy   1441 SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP 1500
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP 1500

Qy   1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560

Qy   1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTTYIYRNNTQIAS 1620
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTTYIYRNNTQIAS 1620

Qy   1621 GVTETTYRDPDLIATGFYTYGVKVVYPNGESAIETATNLNITSLADVTAQKPYTLTVVGKTI 1680
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GVTETTYRDPDLIATGFYTYGVKVVYPNGESAIETATNLNITSLADVTAQKPYTLTVVGKTI 1680

Qy   1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYEKLAVK 1732
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYEKLAVK 1732


RESULT 2
US-08-353-485-10
; Sequence 10, Application US/08353485
; Patent No.5830710
;   GENERAL INFORMATION:
;      APPLICANT: Progulske-Fox, Ann
;      APPLICANT: Tunwasorn, Somying
;      APPLICANT: Lepine, Guylaine
;      APPLICANT: Han, Naiming
;      APPLICANT: Lantz, Marilyn
;      APPLICANT: Patti, Joseph
;      TITLE OF INVENTION:  Cloned Porphyromonas gingivalis Genes
;      TITLE OF INVENTION:  and Probes for the Detection of Periodontal Disease
;      NUMBER OF SEQUENCES: 12
;      CORRESPONDENCE ADDRESS:
;         ADDRESSEE: Ted W. Whitlock
;         STREET: 2421 N.W. 41st Street, Suite A-1
;         CITY: Gainesville
;         STATE: FL
```

```
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/353,485
    FILING DATE: 09-DEC-1994
    CLASSIFICATION: 424
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/647,119
    FILING DATE: 25-JAN-1991
    CLASSIFICATION: 424
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/241,640
    FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
    NAME: Whitlock, Ted W.
    REGISTRATION NUMBER: 36,965
    REFERENCE/DOCKET NUMBER: UF15.C2
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (904) 375-8100
    TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 10:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 1732 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-353-485-10

Query Match         100.0%;  Score 9179;  DB 2;  Length 1732;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 1732;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy   1  MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCNNSFKQFDASFSFNEVELTKVETK   60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1  MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCNNSFKQFDASFSFNEVELTKVETK   60

Qy  61  GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVVSLNQYGSEKL  120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61  GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVVSLNQYGSEKL  120

Qy 121  MPHQPSMSKSDDPEKVPFVVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD  180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121  MPHQPSMSKSDDPEKVPFVVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD  180

Qy 181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT  240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT  240

Qy 241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA  300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA  300

Qy 301  ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301  ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360

Qy 361  IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNY   420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361  IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYNQEHGYTDVYNY   420

Qy 421  LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNDKYFLAIGNCCITA   480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421  LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNDKYFLAIGNCCITA   480

Qy 481  QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS  540
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481  QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS  540

Qy 541  YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP  600
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541  YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP  600

Qy 601  KINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG  660
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601  KINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG  660

Qy 661  NYDVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQQKVTLKWEAPSAKKAEGSREVK   720
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 661  NYDVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQQKVTLKWEAPSAKKAEGSREVK   720

Qy 721  RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT   780
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721  RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT   780

Qy 781  GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI  840
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781  GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI  840

Qy 841  AGDGGNQPARYDDFTEEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE  900
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841  AGDGGNQPARYDDFTEEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE  900

Qy 901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV  960
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV  960

Qy 961  TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY    1020
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 961  TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY    1020

Qy 1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASS  1080
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASS  1080

Qy 1081 TGNDASNFTNALLEETTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM   1140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 TGNDASNFTNALLEETTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM   1140

Qy 1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA  1200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA  1200

Qy 1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA   1260
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA   1260

Qy 1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL  1320
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL  1320

Qy 1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT  1380
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT  1380

Qy 1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDALLKWNAPASKRAEVLNEDFENGIPA  1440
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDALLKWNAPASKRAEVLNEDFENGIPA  1440

Qy 1441 SWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVSSASHINFEGPQNPDNVLVTPELSLP   1500
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 SWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVSSASHINFEGPQNPDNVLVTPELSLP   1500

Qy 1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR  1560
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR  1560

Qy 1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYIYRNNTQIAS   1620
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYIYRNNTQIAS   1620

Qy 1621 GVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI  1680
```

```
RESULT 3
US-09-066-330-11
; Sequence 11, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275
; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1732
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-09-066-330-11

Query Match      99.9%;   Score 9167;   DB 4;   Length 1732;
Best Local Similarity   99.8%;   Pred. No. 0;
Matches 1729;   Conservative   3;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     1 MRKLLLLIAASLLGVLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MRKLLLLIAASLLGVLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK 60

Qy    61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVVSLNQYGSEKL 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVVSLNQYGSEKL 120

Qy   121 MPHQPSMSKSDDPEKVPFVVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 MPHQPSMSKSDDPEKVPFVVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180

Qy   181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240

Qy   241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300

Qy   301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPELTNI 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPELTNI 360

Qy   361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNY 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNY 420

Qy   421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQKALTNKDKYFLAIGNCCITA 480
           ||||||||||||||||||||||||||||||||||||||| :|||||||||||||||||||
Db   421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480

Qy   481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS 540

Qy   541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP 600

Qy   601 KTNYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 KTNYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660

Qy   661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTGQGKVTLKWEAPSAKKAEGSREVK 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTGQGKVTLKWEAPSAKKAEGSREVK 720

Qy   721 RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Qy   781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEASGKMWI 840
           |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db   781 GTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEASGKMWI 840

Qy   841 AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900

Qy   901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKV 960

Qy   961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY 1020

Qy  1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080

Qy  1081 TGNDASNFTNALLEETTTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVARHFQSTDM 1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 TGNDASNFTNALLEETTTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVARHFQSTDM 1140

Qy  1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200

Qy  1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260

Qy  1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320

Qy  1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380
           ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380

Qy  1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
           |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 AGVSPKKCVNVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440

Qy  1441 SWKTIDADGDGNNWTTPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP 1500
           ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db  1441 SWKTIDADGDGNNWTTPPGGSSFAGHNSAICVSSASYINFEGPQNPDNYLVTPELSLP 1500

Qy  1501 GGGLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
           |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560

Qy  1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIAS 1620
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIAS 1620

Qy  1621 GVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYLTVVGKTI 1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 GVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYLTVVGKTI 1680

Qy  1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
```

```
Db   1621 GVTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI 1680

Qy   1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGHYAVMVVDGKSYVEKLAVK 1732
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGHYAVMVVDGKSYVEKLAVK 1732


RESULT 4
US-08-570-311-27
; Sequence 27, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT:  Progulske-Fox, Ann
;   APPLICANT:  Tumwasorn, Somying
;   APPLICANT:  Lepine, Guylaine
;   APPLICANT:  Han, Naiming
;   APPLICANT:  Lantz, Marilyn
;   APPLICANT:  Patti, Joseph
;   TITLE OF INVENTION:  Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION:  and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES:  29
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Ted W. Whitlock
;     STREET:  2421 N.W. 41st Street, Suite A-1
;     CITY:  Gainesville
;     STATE:  FL
;     COUNTRY:  USA
;     ZIP:  32606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/570,311
;     FILING DATE:
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/353,485
;     FILING DATE:  09-DEC-1994
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/647,119
;     FILING DATE:  25-JAN-1991
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/241,640
;     FILING DATE:  08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Whitlock, Ted W.
;     REGISTRATION NUMBER:  36,965
;     REFERENCE/DOCKET NUMBER:  UF15.C3
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (904) 375-8100
;     TELEFAX:  (904) 372-5800
;   INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1358 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;       MOLECULE TYPE:  protein
US-08-570-311-27

  Query Match         67.0%;  Score 6151.5;  DB 2;  Length 1358;
  Best Local Similarity  85.0%;  Pred. No. 0;
  Matches 1170;  Conservative  60;  Mismatches  118;  Indels  29;  Gaps  10;

Qy    366 MYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVVNYLKAPY 425
          |||||||||||||||||||||||| :::||:||||::||:::||||:||:|
Db      1 MYEKATMPDKSYLEKALLIAGADSYWNPKIGQQTIKYAVQYYYNQDHGYTDVVSYPKAPY 60

Qy    426 TGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYV 485
          |||||||||||||||||||||||||||    |:|||||||||||||||||||||||||
Db   1138 E--GAKKTYKVGSPGLPQWSHDHYALCISKSGTAAADF------EVIFEETMTYTQGGA 1189
```

```
Qy  1557  RGTRAQGTWYQKTVQLPAGTKYAFRHFGCTDFFWINLDDVVIT-SGNAPSYTYTIYRNN 1615
          ||.|:                  ||||| :|: |:.::|   :
Db  1190  NLTREK------DLPAGTKYVAFRHYNCTDVLGIMIDDVVITGEGEGPSYTYTVYRDG 1241

Qy  1616  TQIASGVTETTYRPDLATGFYTGYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTV 1675
          :.:|||||| ||||    |:  |    ||::::|       :|
Db  1242  TKIQEGLIETTYRDAGMSAQSHEYCVEVKIAAGVSPKVCVDYIPDGVADVTAQKPYTLTV 1301

Qy  1676  VGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
          ||||||||||||||||||||||||||||||||||||:|||||||||||||||||:|
Db  1302  VGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGYYAVMVVVDGKSYVEKLAIK 1358
```

RESULT 5
US-08-141-324-14
; Sequence 14, Application US/08141324
; Patent No. 5475097
; GENERAL INFORMATION:
;   APPLICANT: Travis, James
;   APPLICANT: Potempa, Jan S.
;   APPLICANT: Barr, Philip J.
;   APPLICANT: Pavloff, Nadine
;   APPLICANT: Pike, Robert N.
;   TITLE OF INVENTION: Lysine-specific Porphyromonas gingivalis
;   TITLE OF INVENTION: Protease
;   NUMBER OF SEQUENCES: 28
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenlee and Winner, P.C.
;   STREET: 5370 Manhattan Circle, Suite 201
;   CITY: Boulder
;   STATE: CO
;   COUNTRY: US
;   ZIP: 80303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/141,324
;   FILING DATE: 21-OCT-1993
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ferber, Donna M.
;   REGISTRATION NUMBER: 33,878
;   REFERENCE/DOCKET NUMBER: 44-93
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 303-499-8080
;   TELEFAX: 303-499-8089
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 942 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-141-324-14

Query Match   52.4%; Score 4811; DB 1; Length 942;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 911; Conservative 17; Mismatches 14; Indels 0; Gaps 0;

Qy   1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTRRTCTNNSFKQFDASFSNEVELITKVETK 60
         |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db   1 MRKLLLIAASLLGVGLYAQNAKIKLDAPTRRTCTNNSFKQFDASFSNEVELITKVETK 60

Qy   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVVSLNQYGSEKL 120
         ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVYSLNQYGSEKL 120

Qy 121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
         ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db 121 MPHQPSMSKSDDPEKVPFAYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
```

```
Qy  181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240

Qy  241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300

Qy  301  ASAAPVFLAIVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  ASAAPVFLAIVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360

Qy  361  IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQBHGYTDVYNY 420
         ||||:|||||||||||||||||||||||| :.:||:||||||||||||||:.:|||:||
Db  361  IDKVLMYEKATMPDKSYLEKALLIAGADSYWNPKIGQQTIKYAVQYYYNQDHGYTDVYSY 420

Qy  421  LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTSQLKALTNKDKYFLAIGNCCITA 480
         ||.|||||||||||||||||||||||||:||||.:||:||:|||||||||||||.|:
Db  421  PKAPYTGCYSHLNTGVGFANYTAHGSETSWADPSVTATQVKALTNKNKYFLAIGNCCVTA 480

Qy  481  QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFRGTSMGS 540
         ||||:.||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db  481  QFDYPQPCFGEVMTRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFRGTSMGS 540

Qy  541  YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNTHIGAHYYWEAYHVLGDGSVMPYRAMP 600
         ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db  541  YDATFLEDSYNTVNSIMWAGNLAATHAENIGNVTHIGAHYYWEAYHVLGDGSVMPYRAMP 600

Qy  601  KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db  601  KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVNMTKQITENG 660

Qy  661  NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720
         |||||||||||||||||||:||||||||||||||||||||||||||:|||||||||||||
Db  661  NYDVVITRSNYLPVIKQIQAGEPSPYQPVSNLTATTQGQKVTLKWDAPSAKKAEGSREVK 720

Qy  721  RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721  RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Qy  781  GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI 840
         ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db  781  GTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI 840

Qy  841  AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841  AGDGGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE 900

Qy  901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGS 942
         ||||||||||||||||||||||||||||||||||||||||||
Db  901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGS 942
```

RESULT 6
US-08-541-902-14
; Sequence 14, Application US/08541902
; Patent No. 5707620
; GENERAL INFORMATION:
;   APPLICANT: Travis, James
;   APPLICANT: Potempa, Jan S.
;   APPLICANT: Barr, Philip J.
;   APPLICANT: Pavloff, Nadine
;   APPLICANT: Pike, Robert N.
;   TITLE OF INVENTION: Lysine-specific Porphyromonas gingivalis
;   TITLE OF INVENTION: Protease
;   NUMBER OF SEQUENCES: 28
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenlee and Winner, P.C.
;   STREET: 5370 Manhattan Circle, Suite 201
;   CITY: Boulder
;   STATE: CO
;   COUNTRY: US

```
      ZIP: 80303
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/541,902
   FILING DATE:
   CLASSIFICATION: 435
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US/08/141,324
   FILING DATE: 21-OCT-1993
ATTORNEY/AGENT INFORMATION:
   NAME: Ferber, Donna M.
   REGISTRATION NUMBER: 33,878
   REFERENCE/DOCKET NUMBER: 44-93
TELECOMMUNICATION INFORMATION:
   TELEPHONE: 303-499-8080
   TELEFAX: 303-499-8089
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
   LENGTH: 942 amino acids
   TYPE: amino acid
   TOPOLOGY: linear
   MOLECULE TYPE: protein
US-08-541-902-14

Query Match             52.4%;    Score 4811;    DB 1;    Length 942;
Best Local Similarity   96.7%;    Pred. No. 0;
Matches 911; Conservative 17; Mismatches 14; Indels 0; Gaps 0;
```

```
Qy      1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK  60
            |||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db      1 MRKLLLIAASLLGVGLYAQNAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK  60

Qy     61 GGTFASVSIPGAFFTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVVSLNQYGSEKL 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGTFASVSIPGAFFTGEVGSPEVPAVRKLIAVPVGATPVVRVKSFTEQVVSLNQYGSEKL 120

Qy    121 MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 MPHQPSMSKSDDPEKVPFAYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD 180

Qy    181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240

Qy    241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300

Qy    301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360

Qy    361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVVNY 420
            |||||||||||||||||||||:||||||:::|||||::||||:||||||||||||:::||
Db    361 IDKVLMYEKATMPDKSYLEKALLIAGADSYWNPKIGQQTIKYAVQYYYNQDHGYTDVVSY 420

Qy    421 LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480
            ||||||:|||||||||||||||||||:|||||||||:::||:|||||||||||||||||:|
Db    421 PKAPYTGCYSHLNTGVGFANYTAHGSETSWADPSVTATQVKALTNKNKYFLAIGNCCVTA 480

Qy    481 QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYWSGANAVFGVQPTFEGTSMGS 540
            ||||:||||||:|||:||||||||||||||||||||||||||||||||||||||||||
Db    481 QFDYPQPCFGEVMTRVKEKGAYAYIGSSPNSYWGEDYWSGANAVFGVQPTFEGTSMGS 540

Qy    541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMP 600
            ||||||||||||||||||||||||||:|||||:||||||||||||||||||||||||||
Db    541 YDATFLEDSYNTVNSIMWAGNLAATHAENIGNVTHIGAHYYWEAYHVLGDGSVMPYRAMP 600

Qy    601 KINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVNMTKQITENG 660
            |||||||||||||||||||||:||||||||||||||||||||||||:||||||||||||
Db    601 KINTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVNMTKQITENG 660

Qy    661 NYDVVITRSNYLPVIKQICIQVGBPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK 720
            ||||||||||||||||||||||:||||||||||||||||||||||||||:|||||||||||
Db    661 NYDVVITRSNYLPVIKQICIQAGEPSPYQPVSNLTATTQGQKVTLKWDAPSAKKAEGSREVK 720

Qy    721 RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT 780

Qy    781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI 840
            |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db    781 GTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWI 840

Qy    841 AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTMEVEDDSPASYTYTVVRDGTKIKE 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTMEVEDDSPASYTYTVVRDGTKIKE 900

Qy    901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGS 942
            |||||||||||||||||||||||||||||||||||||||||
Db    901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGS 942
```

```
RESULT 7
US-08-570-311-8
; Sequence 8, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tunwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Ted W. Whitlock
;     STREET: 2421 N.W. 41st Street, Suite A-1
;     CITY: Gainesville
;     STATE: FL
;     COUNTRY: USA
;     ZIP: 32606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/570,311
;     FILING DATE:
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/353,485
;     FILING DATE: 09-DEC-1994
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/647,119
;     FILING DATE: 25-JAN-1991
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/241,640
;     FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Whitlock, Ted W.
;     REGISTRATION NUMBER: 36,965
;     REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (904) 375-8100
;     TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
```

```
Db    897  GTAAADF------EVIFEETMTYTQGGANLTREK-------DLPAGTKYVAFRHYNCTD 942
                                                        |||||||||||||||||
Qy   1589  FFWINLDDVVIT-SGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVVPN 1647
                  :|:||    :|| :                     | |::|||:||
Db    943  VLGIMIDDVVITGEGEGPSYTYVYRDGTKIQEGLTETTYRDAGMSAQSHEYCVEVKYAA 1002

Qy   1648  GESAIETATLNITSLADVTAQKPYTLTVVGKTITVTVCCQGEAMIYDMNGRRLAAGRNTVVY 1707
                                                 ||||||||||||||||||||||||
Db   1003  GVSPKVCVDYIPDGVADVTAQKPYTLTVVGKIITVTCCQGEAMIYDMNGRRLAAGRNTVVY 1062

Qy   1708  TAQGGHYAVMVVVDGKSYVEKLAVK 1732
                   ||||||||||||||||||:|
Db   1063  TAQGGYYAVMVVVDGKSYVEKLAIK 1087
```

RESULT 8
US-08-353-485-8
; Sequence 8, Application US/08353485
; Patent No. 5830710
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naining
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
;   INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1087 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-353-485-8

```
Qy   638  GTGVANASGVATVSMKQITENGNYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQ 697
               :|:::||                |||||||||||||||||||||||||||||
```

; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1087 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-8

Query Match   50.9%; Score 4673.5; DB 2; Length 1087;
Best Local Similarity 81.4%; Pred. No. 1e-312;
Matches 900; Conservative 56; Mismatches 120; Indels 29; Gaps 10;

```
Qy    638  GTGVANASGVATVSMKQITENGNYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQ 697
                :|:::||                |||||||||||||||||||||||||||||
Db      2  GTVVADPTVAAPVKMAKQIAENGNYDVVMTRSNYLPVINQIQAGEPSPYQPVNNLTAPPE 61

Qy    698  GQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGY 757
                 :|:|||   ||| |:        ||||||:|||||||||||||||||||||||||
Db     62  GEEVALKWDTPSAKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGY 121

Qy    758  QFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEV 817
                 ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db    122  QFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEV 181

Qy    818  VIPGGVVDYCITNPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDM 877
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182  VIPGGVVDYCITNPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDM 241

Qy    878  EVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDV 937
                |||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db    242  EVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDV 301

Qy    938  TVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPNGTTLSESFENGIPASW 997
                |||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db    302  TVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPNGT-TTLSESFENGIPASW 360

Qy    998  KTIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGK 1057
                |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db    361  KTIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLANGGK 420

Qy   1058  LTFWVCAQDANYASEHYAVNDGFFGPGDHYAVMISKTGNAGDFTVVFETPNGINKGGARFGLSTEANGAKPQSVWI 1117
                ||||||||||||                      |||:||:||||||:||||||||||||||||||||||||
Db    421  LTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETIITAKGVRSPKAIRGRIQCTWR 480

Qy   1118  QKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTT 1177
                |||||||||||||||||:|:||||||||||||||||||||||||||||||||||||||
Db    481  QKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTT 540

Qy   1178  IDADGDGQGWLCLSSGGLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYY 1237
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  IDADGDGQGWLCLSSGGLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYY 600

Qy   1238  YAVNDGFFGPGDHYAVMISKTGNAGDFTVVFETPNGINKGGARFGLSTEANGAKPQSVWI 1297
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  YAVNDGFFGPGDHYAVMISKTGNAGDFTVVFETPNGINKGGARFGLSTEANGAKPQSVWI 660

Qy   1298  ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLT 1357
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  ERTVDLPAGTKYVAFRHYNCSDLDYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLT 720

Qy   1358  ETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAIL 1417
                ||||||||||||||||||||||||||||||||||||||||||||||||||:
Db    721  ETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTINPTQFNPVKNLKAQ--PDGGDVVL 778

Qy   1418  KWNAPASKRAEVLNEDFE-NGIPASWKTIDADGDGNNW--TTTPPGG-----SSFAGHN 1469
                ||:|   |::        | :  :|   |  |  |  |            :
Db    779  KWEAPSGKRGELLNEDFEDGDAIPTGWTALDADGDGNNWDITLNEFTRGERHVLSPLRASN 838

Qy   1470  SAICVSSASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQD-ANYASEHYAVVASST 1528
                :|| :                  : ::|: | |::    |:  |  :::  :
Db    839  VAISYSSLLQQOEYLPLTPNNFLITPKVE--GAKKITYKVGSPGLPQWSHDHYALCISKS 896

Qy   1529  GNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTD 1588
                |||             :   :                  |||||||||||||||||
```

RESULT 9
US-09-066-330-10
; Sequence 10, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric C.
;   APPLICANT: Bhogal, Peter S.
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;   FILE REFERENCE: Reynolds
;   CURRENT APPLICATION NUMBER: US/09/066,330A
;   CURRENT FILING DATE: 1998-09-15
;   EARLIER APPLICATION NUMBER: PN 6275
;   EARLIER FILING DATE: 1995-10-30
;   EARLIER APPLICATION NUMBER: PCT/AU96/00673
;   EARLIER FILING DATE: 1996-10-30
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 1706
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-09-066-330-10

  Query Match   46.5%; Score 4272; DB 4; Length 1706;
  Best Local Similarity   50.9%; Pred. No. 8.9e-285;
  Matches 932; Conservative 204; Mismatches 456; Indels 240; Gaps 47;

Qy   2 GTVVADPTVAAPVKMAKQIAENGNYDVVMTRSNVLPVINQIQAGEPSPYQPVNNLTAPPE 61
            :::|| :::|||||:||||||||||||||||||||||||||||||||||||||||||||
Db   698 GQKVTLKWEAPSAKKAEGSREYKRIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGY 757

Qy   62 GEEVALKWDTPSAKKAEASREVKRIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGY 121
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   758 QFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEV 817

Qy   122 QFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEV 181
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   818 VIPGGVYDYCIINPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDM 877

Qy   182 VIPGGVYDYCIINPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDM 241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   878 EVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDV 937

Qy   242 EVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDV 301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   938 TVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASW 997

Qy   302 TVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGT-TTLSESFENGIPASW 360
            ||||||||||||||||||||||||||||||||||||||||||| :||||||||||||||
Db   998 KTIDADGDGHGWKPGNAPGIAGVNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGK 1057

Qy   361 KTIDADGDGHGWKPGNAPGIAGVNSNGCVYSESFGLGGIGVLTPDNYLITPALDLANGGK 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   1058 LTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWR 1117

Qy   421 LTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPEAIRGRIQGTWR 480
            ||||||||||||||||||||||||||||||||||||||||||||||| :||||||||||
Db   1118 QKTVDLLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADPTETFESSTHGEAPAEWTT 1177

Qy   481 QKTVDLLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADPTETFESSTHGEAPAEWTT 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1178 IDADGDQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYY 1237

Qy   541 IDADGDQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYY 600
            |||||||||||||||||||||||| :|||:||||||||||||||||||||||||||||
Db   1238 YAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWI 1297

Qy   601 YAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWI 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1298 ERTVDLLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYTVYTVYRDGTKIKEGLT 1357

Qy   661 ERTVDLLPAGTKYVAFRHYNCSDLDYILLDDIQFTMGGSPTPDYTVYTVYRDGTKIKEGLT 720
            |||||||||||||||||||||||| :|||||||||||||||||||||||||||||||||
Db   1358 ETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAIL 1417

Qy   721 ETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTINPTQFNPVKNLKAQ--PDGGDVVL 778
            ||||||||||||||||||||||||||||| :||: ||:||||:|||:|| |||| ||
Db   1418 KWNAPASKRAEVLNEDFE-NGIPASWKTIDADGDGNNW--TTTPPGG-----SSPAGHN 1469

Qy   779 KWEAFSGKRGELLNEDFEGDAIPTGWTALDADGDGNNWDIILNEFTRGERHVLSPLRASN 838
            || :||| :|||||| |:|| |||: |:||||||||| : | |:||| | |:|
Db   1470 SAICVSSASHINFEGPQNPDNVLVTPELSLPGGGTLTFWVCAQD-ANYASEHYAVVASST 1528

Qy   839 VAISYSSLLQQOEYLPLTPNNFLITPKVE--GAKKITYKVGSPGLPQWSHDHYALCISKS 896
            ||:|:|| ||| ||| | |||||||||:| ||||||||||||| || |||||| |:||
Db   1529 GNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTD 1588

Qy   897 GTAAADF------EVIFEETMTYTQGGANLTREK------DLPAGTKYVAFRHNCTD 942
            :||| | :|:||||:| ::||||||:|:   ::|||||||||||:||
Db   1589 FFWINLDDVVIT-SGNAPSYTYTYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVPN 1647

Qy   943 VLGIMIDDVVITGEGEGPSYTYTVYRDGTKIQEGLTETTYRDAGMSAQSHEYCVEVKYAA 1002
            |:| :||||| |:||: |:|||||||| || ||| ||||||| : |:|||||||| ||
Db   1648 GESAIETATLNITSLADVTAQKPYTLTVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVY 1707

Qy   1003 GVSPKVCVDYIPDGVADVTAQKPYTLTVGKTITVTCCQGEAMIYDMNGRRLAAGRNTVVY 1062
            ||| :|| |:||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1708 TAQGHYAVMVVDGKSYVEKLAVK 1732

Qy   1063 TAQGGYYAVMVVDGKSYVEKLAIK 1087

```
Qy 1641 VKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTITVTCCQGEAMIYDMNGRRLAA 1700
             ||:|       |||||||||||||||||||||||||||||||||||
Db 1615 VEVKYTAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCCQGEAMIYDMNGRRLAA 1674

Qy 1701 GRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
        ||||||||||||:||||||||||||||||:|:
Db 1675 GRNTVVYTAQGGYYAVMVVVDGKSYVEKLAIK 1706
```

RESULT 10
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
GENERAL INFORMATION:
;   APPLICANT: Travis, James
;   APPLICANT: Potempa, Jan S.
;   APPLICANT: Barr, Philip J.
;   APPLICANT: Pavloff, Nadine
;   TITLE OF INVENTION: Porphyromonas gingivalis
;   TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
;   STREET: 5370 Manhattan Circle, Suite 201
;   CITY: Boulder
;   STATE: Colorado
;   COUNTRY: US
;   ZIP: 80303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/336,308A
;   FILING DATE: 08-NOV-1994
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/119,361
;   FILING DATE: 10-SEP-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/265,441
;   FILING DATE: 24-JUN-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ferber, Donna M.
;   REGISTRATION NUMBER: 33,878
;   REFERENCE/DOCKET NUMBER: 21-93C
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (303) 499-8080
;   TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1704 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-336-308A-10

```
Query Match             46.5%;  Score 4270;  DB 3;  Length 1704;
Best Local Similarity   50.9%;  Pred. No. 1.2e-284;
Matches  932;  Conservative  204;  Mismatches  456;  Indels  238;  Gaps  47;

Qy   6 LLIAASLLGVLGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSNEVELTKVETKGGTFA 65
         :|  :|| ::|||||  :    :  |   ::|:.|  : |   :.:|:::|| |||.|
Db  10 IALCSSLLGGMAFAQQTELGIRGNPNVRLLESTQGSVTK--VQFRMDNLKFTEVQTPKGI-- 65

Qy  66 SVSIPGAFPT------GEVGSPEVPAVRKLLAVPVGATPVVRVKS--FTEQVYSLNQYG 116
                           ||  .| .|     |:|:  |   :|.|:
Db  66 ----GQVPTYTBGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL---- 116

Qy 117 SEKLMPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINP 176
              |::         |:|: | ::.  :    ||  : ||   | :: | :|| |:
Db 117 ---IAPSKGMIMRNEDPKKIPYYY-GKSYSQNKFFPGEIATLDDPFIIRDVRGQVVNFAP 172
```

```
Qy  612 PQNQASYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMVKQITENGNYDVVITRSN 670
            ::   :  :|:  :|||||||        :  :  : |    :|::        : :
Db  590 NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGYN 647

Qy  671 YLPVIKQIQV-GEPSPYQPVSNLTATTGGQKVTLKWEAPSAK---KAEGSREVKRIGDGL 726
          |||| : :  ||:|:|||||||||||||||||||||:|:||     |    :|: :|:
Db  648 KETVIKTINTNGEPNPYQPVSNLTATTGGQKVTLKWDAPSTKTNATTNTARSVDGIRELV 707

Qy  727 FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA 783
         :.:  .|  ::  :  :||  ::||.|.:|||:||||||: .|.|||  :  : ||  :
Db  708 LLSVSDAPELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQVGQVIPSDTHTIWPNCSV 767

Qy  784 SSNLYSANFEYLVPANADPVVTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGD 843
             |  ||:||:|| ||||   ::::|.||  ||||||.|:| | ||||||||||:|||
Db  768 PANLF-APFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQ 825

Qy  844 GGNQPARYDFFTEAGKKYTFTMRRAGMGDGTMEVEDDSPASYTIVYRDGTKIKEGLT 903
        ||:||| ||:| ||||| |::::: ||||||: :    || ||| ||||||||||||
Db  826 G---PTKEDDYVFEAGKKYHFLMKKMGSGDGTELTISRGGGSDYTYTVVYRDGTKIKEGLT 882

Qy  904 ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK 963
        ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db  883 ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK 942

Qy  964 WDAPNGTPNPNPNPNPNPG---TTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020
        |||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db  943 WDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1002

Qy 1021 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASS 1080
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1003 NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASS 1062

Qy 1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
        ||||||||||||||||||||||||||:|:||||||||||||||||||||||||||||||||
Db 1063 TGNDASNFTNALLEETITAKGVRSPEAMRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1122

Qy 1141 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1123 FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1182

Qy 1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGTNA 1260
        |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1183 HGGTNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGTNA 1242

Qy 1261 GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
        ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db 1243 GDFTVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1302

Qy 1321 NYILLDDIQFTMGGSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1303 NYILLDDIQFTMGGSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1362

Qy 1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE----------- 1428
        ||||||||.|||||||||||||||  :   |     ||:|||:|:
Db 1363 AGVSPKKCVNVTVNSTQFNPVQNLKAQ--PDGGDVVLKWEAPSAKKTBGSREVKRIGDGL 1420

Qy 1429 ---------------VLNEDFENGIPASWKTIDADGDGNWTTPPPGGGSSFAGHNSA 1471
                       || :|:|||||||||||||||||.  ||| |||:::|||::|
Db 1421 FVTIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASS 1479

Qy 1472 ICVSSASHINFEG--PQNPD-------NYLVT--PELSLPGGGTLTFWVCAQDANYASEHY 1521
         : :|: ||::|   |::|          ::|  :|:|:|||| | ||:| ||||||:|:|
Db 1480 DLYSA-----NFESLIPANADPVTTQNIIVTGQGEVVIPGG---VIDYCITNPEPASGKM 1532

Qy 1522 AVVASSTGNDASNFANALLEEVLTAKTVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-A 1580
                    |        :|     :|     |:|:|| |    ||:||||||| |
Db 1533 WI-AGDGGNQPARYDDFTFE------------------------------AGKKYTFT 1559

Qy 1581 FRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLLATGFYTYG 1640
        :|||| ||:| :|||:|:||| ||||| ||:||:||:|||| ||||||| || |: ||:| 
Db 1560 MRRAGMGDG-----TDMEVEDDSPASYTIVYRDGTKIKEGLTETTYRDAGMSAQSHEYC 1614
```

```
Qy   177  VQYDVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVVTD  233
Db   173  LQYNPVTKLRIYTEITVAVSETSEQGKNILNKKGTFAG-----FEDTYKRMF-----MN  222

Qy   234  HGDLYNTPV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASI  287
Db   223  YEPGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI  282

Qy   288  KAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYYSAVDGDYF  341
Db   283  QQFVKQEYEKEGND------LTYVLLVGDHKDIPAKITPGIKSDQV---YGQIVGNDHY   332

Qy   342  PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK  401
Db   333  NEVFIGRFSCESKEDLKTQIDRTTHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQ  392

Qy   402  Y-GMQYYINQBHGYTDVINYLKAPYTGCY---------SHLNTGVSFANYTAHGSETAW   450
Db   393  HENVIANLLTQYGYTKIIK--------CYDPGVTPKNIIDAFNGGISLVNYTGHGSETAW  444

Qy   451  ADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV---KEKGAYAYIG   506
Db   445  GTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIA  504

Qy   507  SSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDAIFLEDSYNTVNSIMWAGNLAATH  566
Db   505  ASPMRG------------QDEMNEI-----------LCEKH   529

Qy   567  AGNI----GNITHIGAHYYWEAYH----------VLGDGSVMPYRAMPKTNTYTLPASL   611
Db   530  PNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQI  589

Qy   612  PQNQASYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN  670
Db   590  NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLILTVVGYN  647

Qy   671  YLPVIKQIQV-GEPSPYQPVSNLTATTGSQKVTLKWEAPSAK---KABGSREVKRIGDGL  726
Db   648  KETVIKTINTNGENPYQPVSNLTATTQQGKVTLKWDAPSTKTNATTNTARSVDGIRELV   707

Qy   727  FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA  783
Db   708  LLSVSDAPELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQYGQVIFSDTHTLWPNCSV  767

Qy   784  SSNLYSANFEYLVPANADPVVTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGD   843
Db   768  PANLF-APFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAFQ-ANAKIWIAGQ  825

Qy   844  GGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLT  903
Db   826  G---PTKEDDYVFBAGKKYHFLMKKMGSGDGTELTISEGGSDYTYTVVRDGTKIKEGLT   882

Qy   904  ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK  963
Db   883  ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK  942

Qy   964  WDAPNGTPNPNPNPNPNPG-TTLSESFENGIPASWKTIIDADGDGHGWKPGNAPGIAGYNS 1022
Db   943  WDAPNGTPNPNPNPNPNPGTTTLSESFENGIPASWKTIIDADGDGHGWKPGNAPGIAGYNS 1002

Qy  1023  NGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG 1082
Db  1003  NGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKITFWVCAQDANYASEHYAVVASSTG 1062

Qy  1083  NDASNFTNALLEETTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHPQSTDMFY  1142
Db  1063  NDASNFTNALLEETTAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHPQSTDMFY  1122

Qy  1143  IDLLDEVEIKANGKKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHG 1202
Db  1123  IDLLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHG 1182
```

```
Qy  1203  GSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGD 1262
Db  1183  GTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGD 1242

Qy  1263  FTVVFEETPNGINKGGARFGLSTRANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNY 1322
Db  1243  FTVVFEETPNGINKGGARFGLSTRANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNY 1302

Qy  1323  ILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFBEDGVATGNHEYCVEVKYTAG 1382
Db  1303  ILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFBEDGVATGNHEYCVEVKYTAG 1362

Qy  1383  VSPKKCVDVTVNSTQFNPVQNLTABQAPNSMDAILKWNAPASKRAE-------------- 1428
Db  1363  VSPKECVNVTINPTQFNPVXNLKAQ--PDGGDVVLKWEAPSAKKTEGSREVKRIGDGLFV 1420

Qy  1429  ---------------VLNEDFENGIPASWKTIDADGDGNWTTPPPGGSSFAGHNSAIC  1473
Db  1421  TIEPANDVRANEAKVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSNL 1479

Qy  1474  VSSASHINFBG--PQNPD------NYLVT-PELSLPGGGTLTFWVCAQDANYASEHYAV 1523
Db  1480  YSA----NFEYLIPANADPVVTQNIVITGQGEVVIPGG---VYDYCITNPEPASGKMWI 1532

Qy  1524  YASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKVV-AFR 1582
Db  1533  -AGDGGNQPARYDDFTFE---------------AGKKYTFTMR  1559

Qy  1583  HFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVK 1642
Db  1560  RAGMGDG-----TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVE 1614

Qy  1643  VVYPNGESAIETATLNITSLADVTAQKPYTLTVGKTITVTCQGEAMIYDMNGRRLAAGR  1702
Db  1615  VKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGR 1674

Qy  1703  NTVVYTAQGGHYAVMVVVDGKSYVEKLAVK  1732
Db  1675  NTVVTAQGGYYAVMVVVDGKSYVEKLAVK   1704
```

RESULT 11

```
;
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 103-95 WO
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (303) 488-8080
;   TELEFAX: (303) 499-8099
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1704 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-822-324-6

Query Match            46.5%;  Score  4270;  DB 3;  Length 1704;
Best Local Similarity  50.9%;  Pred. No. 1.2e-284;
Matches  932;  Conservative  204;  Mismatches  456;  Indels  238;  Gaps  47;
```

```
Qy    6 LLIAASLLGVGLYAQSAKIKLDAPTIRTTCINNSFKQFDASFSFNEVELTKVETKGGTFA 65
          ||  ||      :: :|            |       :: |::|:|
Db   10 IALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGI-- 65

Qy   66 SVSIPGAFPT------------GEVGSPEVPAVRKLIAVPVGATPVVRVKS--FTEQVYSLNQYG 116
          :      |||            |  : ||| |        ||    :
Db   66 ----GQVPTTVEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL---- 116

Qy  117 SEKLMPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINP 176
          :| ::   :      :::|:|:|||           ::
Db  117 ---IAPSKGMIMRNEDDPKKIPVYY-GKSYSQNKFFPGEIATLDDPFILRDVRGQVVNFAP 172

Qy  177 VQYDVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTD 233
          :|:  :                 |||
Db  173 LQYNPVTKTLRIYTEITVAVSSTSEQGKNILNKKGTFAG------FEDTYKRMF-----MN 222

Qy  234 HGDLYNTPV------RMLVVAGAKFKEALKPWLTWKAQKGFVLDVHYTDEARVGTNASI 287
          :  ::|||      ||||
Db  223 YEPGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI 282

Qy  288 KAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYSAVDGDYF 341
          :: :::|         ||||
Db  283 QQFVKQEYEKEGND------LTYVLLVGDHKDIPAKITPGIKSDQV--YGQIVGNDHY 332

Qy  342 PEMYTFRMSASSPEELTNIIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK 401
          :: ::|||||
Db  333 NEVFIGRFSCESKEDLKTQIDRTIHYRNITTEDKWIGQALCIASAGGPSADNGESDIQ 392

Qy  402 Y-GMQYYYNQEHGYTDVYNYLKAPYTGCY------SHLNTGVSFANYTAHGSETAW 450
          : :      |||||
Db  393 HENVIANLLTQYGYTKIIK------CYDPGVTPKNIIDAFNGGGISLVNYTGHGSETAW 444

Qy  451 ADPLLTTSQLKALTNKDKYFLAIGNCCITAQPFDYVQPCFGEVITRV---KEKGAYAYIG 506
          :  ||||||
Db  445 GTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIA 504

Qy  507 SSPNSYWGEDYMWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH 566
          :  ||||
Db  505 STINQSW-----ASPMRG-------------QDEMNEI-------LCEKH 529

Qy  567 AGNI----GNITHIGAHYYWEAYH------------VLGDGSVMPYRAMPKTNYTLPASL 611
          :: ||||
Db  530 PNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQI 589

Qy  612 PQNQASYSIQAS-AGSVVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN 670
          :: ||||||
Db  590 NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGYN 647

Qy  671 YLPVIKQIQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK---KAEGSREVKRIGDGL 726
          :: ||||||
Db  648 KETVIKTINTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELV 707

Qy  727 FVTIEPAND-VRANEAKVVLAADNWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA 783
          :: ||||||
Db  708 LLSVSDAPELIRSGQAEIVLEAHDVWNDGSGYIQILLDADHDQYGQVIPSDTHTLWPNCSV 767
```

```
Qy  784 SSNLYSANFEYLVPANADPVVTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGD 843
          :: |||||
Db  768 PANLF-APFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQ 825

Qy  844 GGNQPARYDDFTFEAGKKYTFMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLT 903
          | |||||||||||
Db  826 G---PTKEDDYVFEAGKKYHFLMKKMGSGDTELTISEGGSGDYTYTVYRDGTKIKEGLT 882

Qy  904 ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK 963
          |||||||||||||||
Db  883 ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK 942

Qy  964 WDAPNGTPNPNPNPNPNG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNS 1022
          ||||||||||||||||||
Db  943 WDAPNGTPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNS 1002

Qy 1023 NGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG 1082
          ||||||||||||||||||
Db 1003 NGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTG 1062

Qy 1083 NDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY 1142
          ||||||||||||||||||
Db 1063 NDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY 1122

Qy 1143 IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDCQGWLCLSSGQLDWLTAHG 1202
          ||||||||||||||||||
Db 1123 IDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHG 1182

Qy 1203 GSNVVSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGPPGDHYAVMISKTGTNAGD 1262
          ||||||||||||||||||
Db 1183 GTNVVAGFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGPPGDHYAVMISKTGTNAGD 1242

Qy 1263 FTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNY 1322
          ||||||||||||||||||
Db 1243 FTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNY 1302

Qy 1323 ILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG 1382
          ||||||||||||||||||
Db 1303 ILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG 1362

Qy 1383 VSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE-- 1428
          ||||:|:::|:::
Db 1363 VSPKECVNVTINPTQFNPVKNLKAQ--PDGGDVVLKNEAPSAKKTEGSREVKRIGDGLFV 1420

Qy 1429 ------------VLNEDFENGIPASWKTIDADGDGNNWTTTPPPGGSSFAGHNSAIC 1473
                      :: :::
Db 1421 TIEPANDVRANEAKVVLAADNWGNDTGYQFL-LDADHNTFGGSVIPATGPLFTGTASSNL 1479

Qy 1474 VSSASHINFEG--PQNPD------NYLVT--PELSLPGGGTLTFWVCAQDANYASEHYAV 1523
          :::   :
Db 1480 YSA----NFEYLIPANADPVVTQNIIVTGQGEVVIPGG---VYDYCITNPEPASGKMWI 1532

Qy 1524 YASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGRTRAQGTWYQKTVQLPAGTKYV-AFR 1582
          :::
Db 1533 -AGDGGNQPARYDDFTFE-------------------------AGKKYTFTMR 1559

Qy 1583 HFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVK 1642
          :: |||||
Db 1560 RAGMGDG-----TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVE 1614

Qy 1643 VVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGR 1702
          ||||||||||||||||||
Db 1615 VKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGR 1674

Qy 1703 NTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
          ||||||||||||||||||
Db 1675 NTVVYTAQGGYYAVMVVVDGKSYVEKLAVK 1704
```

```
RESULT 12
US-09-490-931-10
; Sequence 10, Application US/09490931
```

```
; Patent No. 6274718
GENERAL INFORMATION:
;      APPLICANT:  Travis, James
;      APPLICANT:  Potempa, Jan S.
;      APPLICANT:  Barr, Philip J.
;      APPLICANT:  Pavloff, Nadine
;      TITLE OF INVENTION:  Porphyromonas gingivalis
;      TITLE OF INVENTION:  Arginine-specific Proteinase Coding Sequences
;      NUMBER OF SEQUENCES:  16
;      CORRESPONDENCE ADDRESS:
;          ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;          STREET:  5370 Manhattan Circle, Suite 201
;          CITY:  Boulder
;          STATE:  Colorado
;          COUNTRY:  US
;          ZIP:  80303
;      COMPUTER READABLE FORM:
;          MEDIUM TYPE:  Floppy disk
;          COMPUTER:  IBM PC compatible
;          OPERATING SYSTEM:  PC-DOS/MS-DOS
;          SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;          APPLICATION NUMBER:  US/09/490,931
;          FILING DATE:
;      CLASSIFICATION:
;      PRIOR APPLICATION DATA:
;          APPLICATION NUMBER:  08/336,308
;          FILING DATE:
;      PRIOR APPLICATION DATA:
;          APPLICATION NUMBER:  US 08/265,441
;          FILING DATE:  24-JUN-1994
;      ATTORNEY/AGENT INFORMATION:
;          NAME:  Ferber, Donna M.
;          REGISTRATION NUMBER:  33,878
;          REFERENCE/DOCKET NUMBER:  21-93C
;      TELECOMMUNICATION INFORMATION:
;          TELEPHONE:  (303) 499-8080
;          TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO: 10:
;      SEQUENCE CHARACTERISTICS:
;          LENGTH:  1704 amino acids
;          TYPE:  amino acid
;          TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-09-490-931-10
```

```
Qy   1383  VSPKKCVDVTVNSTQFNPVQNLTABQAPNSMDAILKWNAPASKRAE-------------  1428
           |||:||:||||:||| |||||:|| |::||| ||||:
Db   1363  VSPKECVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTEGSREVKRIGDGLFV  1420

Qy   1429  ---------VLNEDFENGIPASWKTIDADGDGNNWTTTPPGGSSFAGHNSAIC  1473
                       : :      :::         :|
Db   1421  TIEPANDVRANEAKVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSNL  1479

Qy   1474  VSSASHINFEG--PQNPD------NYLVT--PELSLPGGGTLTFWVCAQDANYASEHYAV  1523
           :: -     |||||         :|  |||
Db   1480  YSA---NFEYLIPANADPVVTQNIIVTGQGEVVIPGG---VYDYCITNPEPASGKMWI  1532

Qy   1524  YASSTGNDASNFANALLEEVLITAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-AFR  1582
                  |            :::
Db   1533  -AGDGGNQPARYDDFTFE------------------------------AGKKYTFTMR  1559

Qy   1583  HFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTTYGVK  1642
                                               :       |||| ||| |||||
Db   1560  RAGMGDG----TDMEVEDDSPASYTVTYVRDGTKIKEGLTETTYRDAGMSAQSHEYCVE  1614

Qy   1643  VVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTIITVTCCQGEAMIYDMNGRRLAAGR  1702
                          :::||||||||||||||||||||||||||||||||||||||||
Db   1615  VKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTIITVTCCQGEAMIYDMNGRRLAAGR  1674

Qy   1703  NTVVYTAQGGHYAVMVVVDGKSYVEKLAVK  1732
           |||||||||||||||||||||||||||||
Db   1675  NTVVYTAQGGYYAVMVVVDGKSYVEKLAVK  1704


RESULT 13
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT:  Progulske-Fox, Ann
;   APPLICANT:  Tumwasorn, Somying
;   APPLICANT:  Lepine, Guylaine
;   APPLICANT:  Han, Naiming
;   APPLICANT:  Lantz, Marilyn
;   APPLICANT:  Patti, Joseph
;   TITLE OF INVENTION:  Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION:  and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES:  29
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Ted W. Whitlock
;     STREET:  2421 N.W. 41st Street, Suite A-1
;     CITY:  Gainesville
;     STATE:  FL
;     COUNTRY:  USA
;     ZIP:  32606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/570,311
;     FILING DATE:
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/353,485
;     FILING DATE:  09-DEC-1994
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/647,119
;     FILING DATE:  25-JAN-1991
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/241,640
;     FILING DATE:  08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Whitlock, Ted W.
;     REGISTRATION NUMBER:  36,965
```

```
;   REFERENCE/DOCKET NUMBER:  UF15.C3
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (904) 375-8100
;     TELEFAX:  (904) 372-5800
; INFORMATION FOR SEQ ID NO:  29:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1687 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-570-311-29

Query Match             46.2%;  Score 4245;  DB 2;  Length 1687;
Best Local Similarity   51.6%;  Pred. No. 6.3e-283;
Matches  920;  Conservative 197;  Mismatches 443;  Indels 224;  Gaps 44;
```

```
Qy    47  FSFNEVELTKVETKGGTPASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRVKS--  104
          :|:||:||:||:|         ||  : |  |   :|  ||: :|      ||
Db    30  FRMDNLKFTEVQTPKGMAQVPTYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSK  89

Qy   105  FTEQVYSLNQYGSEKLMHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTM  164
          :    :          ||:|   :|    : ::|  | :|: |
Db    90  FIEKKNVL------IAPSKGMIMRNEDPKKIPYVY-GKSYSQNKFFPGEIATLDDPFIL  141

Qy   165  RGVRIAALTINPVQYDVVANQLKVRNNIEI---EVSFQGADEVATQRLYDASFSPYFETA  221
          |  :     |  |: :   |||         :|:  :  : |  :| |      :||
Db   142  RDVRGQVVNFAPLQYNPVTKTLRIYTEITTVAVSETSEQKNILNKKGTFAG------FEDT  196

Qy   222  YKQLFNRDVYTDHGDLYNTPV-------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYT  275
          |:|:|                      ::  |  :|| | :| ::  :    :
Db   197  YKKMF-----MNYEPGRYTPVEBKQNGRMIVIAKKYEGDIKDFVDWKNQRGLRTEVKVA  251

Qy   276  DEAEVGTTNASIKAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVT  329
          :  :::                        : :     |:  ||::|  |   |  |
Db   252  EDIASPVTANAIQQFVKQEYEKKEGND-------LTYVLLVGDHKDIPAKITPGIKSDQV-  303

Qy   330  DLYYSAVDGDYFPEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADY  389
          :: :| |    ||   ::::: :  : | :|   | :|    |  :|    ::|||:
Db   304  --YGQIVGNDHYNEVFIGRFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASABG  361

Qy   390  SWNSQVGQPTIKY-GMQYYYNQEHGYTDVYNYLKAPYTGCY----------SHLNTGVSF  438
          : |  :|          ||            |:||||||          ::  || ||:
Db   362  GPSADNGSDIQHENVIANLLTQYGYTKIIK------------CYDPGVTPKNIIDAFNGGISL  413

Qy   439  ANYTAHGSETAWADLPLLTTSQLKALINKDKYPLAIGNCCITAQPDYVQPCFGEVITRV--  496
          | || | |:                 | : :|  ||:|| ||  :|||||  |||::
Db   414  VNYTGHGSETAWGTSHFGTTHVKQLTNSNQLPIFDVACVNGDFLFSMPCFAEALMRAQK  473

Qy   497  --KEKGAYAYIGGSPNSYWGEDYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVN  554
            :|  |  | |   | |      : ||   | |     :||:|||||||| :||:|||
Db   474  DGKPTGIVAIIASTINQSW----------ASPMRG-                         505

Qy   555  SIMWAGNLAATHAGNI---GNITHIGAHYWBAYH---------VLGDGSVMPYRAM  599
          : || |:  |:|          | || || ||          ||||||| |||
Db   506  ------LCEKHPNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLV  558

Qy   600  PKTNTYLPASLPQNQASYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITE  658
          || ||||||:|| |||||:|||  ||||||||||:||||||||||:|||||||:|||||
Db   559  PTKMQVTAPAQINLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTAINLT-GLTN  616

Qy   659  NGNYDVVITRSNYLPVIKQIQV-GEPSPYQPVSNLTATTGQGKVTLKWEAPSAK--KAE  714
          |||||| ||:||||||||:||: |||| |||||||||||||||||||||:||::  |||
Db   617  ESTLLTIVVGYNKETVIKTINTNGEPNYQPVSNLTATTGQGKVTLKWDAFSTKTNATTN  676

Qy   715  GSREVKRIGDGLFVTIEPAND-VRANEAKVVLAADNVWGDNTGYQPLLDADHNTFGSVIP  773
          ||| |:|| | :: || ||| :   ::  |  ||||  |||| | |  | ||  ||:||
Db   677  TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQYGQVIP  736

Qy   774  A-TGPLFTG-TASSNLYSANPEYLVPANADPVVTQNIIVTGQGEVVIPGGVYDYCITNP  831
          : :: |      ::|| |||:||   |:|  ||    ::| |   |||| | || |||
Db   737  SDTHTIWPNCSVPANLF-APFEYTVPENADPSCSPFNMIMDGTASVNIPAGTYDFAIAAP  795

Qy   832  EPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTV  891
```

US-08-570-311-14

Query Match   45.9%; Score 4217; DB 2; Length 2628;
Best Local Similarity   51.0%; Pred. No. 1.1e-280;
Matches 958; Conservative 114; Mismatches 363; Indels 442; Gaps 53;

```
Qy   179 YDVVANQLKVRNNIEIEVSFQ-----GADEVATQRLYDASFSPYFETAYKQLFNRDVYT 232
             :   ::|:     ::      :|:  :  ::| : : :  ::|:|   | ::
Db   871 YTVYRDGTKIKEGL-TETTFEEDGVATGNHEYCVEVKYTAGVSP--KECVNVTV-- 921

Qy   233 DHGDLYNTPVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIH 292
          :   | :     ::        |    | :        :
Db   922 --DPVQFNPVQNLTGSAVGQKVTLK---WDAPNG---TPNPNPNPGTTTLS------ 966

Qy   293 KKYNDGLAAS------------------AAPVFLALVGDTDVISGEKGKKTKKKVTDLYYSAV 336
          : ::|
Db   967 ESFENGIPASWKTIDADGDNNWTTTPPPGGTSFAGHNSAICASSAS--------YINF 1017

Qy   337 DGDYFPEMYTFRMSASSP--------------EELTNIIDKVLMYEKATMPDKS------YLEK 380
          :  :     ::      |  ::     :|: :
Db  1018 EGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLBE 1077

Qy   381 VL----LIAG--ADYSWNSQVGQPTIKYGMQYYNQEHGYTDVY----- 418
          :|   :  :    :   | ::|:
Db  1078 VLIAKTVVTAPEAIRGTRVQGTWYQKTVQ--LPAGTKYVAFRHFGCTDFFWINLDDVEIK 1135

Qy   419 -NYLKAPYTGCYSHLNTGVSFANVT----AHGSETAWADPLLTTSQLKALTNKDKYFLAIG 474
          :  :|   |:  | : ||:|:| :           | :
Db   796 Q-ANAKIWIAGQG---...PTKEDDYVFEAGKKYHFLMKKMGSGDGTELTISEGGSDYTYTV 851

Qy   892 YRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNL 951
         ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   852 YRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNL 911

Qy   952 TGSSVGQKVTLKWDAPNGTPNPNPNPNPG---TTLSESFENGIPASWKTIDADGDGHG 1008
         |||:|||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db   912 TGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHG 971

Qy  1009 WKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDAN 1068
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   972 WKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDAN 1031

Qy  1069 YASEHYAVYASSTGNDASNFTNALLEETTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTK 1128
         |||||||||||||||||||||||||||||:||||||:|||||||||||||||||||||||
Db  1032 YASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTK 1091

Qy  1129 YVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL 1188
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1092 YVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL 1151

Qy  1189 CLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDH 1248
         ||||||||||||||:||||:|||||||||||||||||||||||||||||||||||||||||
Db  1152 CLSSGQLDWLTAHGSTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDH 1211

Qy  1249 YAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTK 1308
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1212 YAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTK 1271

Qy  1309 YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVAT 1368
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1272 YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVAT 1331

Qy  1369 GNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE 1428
         |||||||||||||||||:|| :|||:||:| |||::||:||:   :   |::| |::|::
Db  1332 GNHEYCVEVKYTAGVSPKECVNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTE 1389

Qy  1429 ------------VLNEDFENGIPASWKTIDADGDGNNWTTTPP 1459
                     :  :|   : :: ::: :    ::
Db  1390 GSREVKRIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFL-LDADHNTFGSVIP 1448

Qy  1460 PGGSSFAGHNSAICVSSASHINFEG--PQNPD----NYLVT--PELSLPGGGTLTFWV 1509
          |   |   ::      :    |::       :|:   |:|||
Db  1449 ATGPLFTGTASSNLYSA----NFEYLIPANADPVVTTQNIIVTGCGEVVIPGG---VYDY 1501

Qy  1510 CAQDANYASEHYAVVASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAAQGTWYQKT 1569
          :     ::  :  |
Db  1502 CIITNPEPASGKMWI-AGDGGNQPARYDDFTFE------------------- 1532

Qy  1570 VQLPAGTKV-AFRHFGCTDFFWINLDDVVITSGNAPSYTYTIYRNNTQIASGVTETTYR 1628
             :    |  :|:    : ::  ::| |||||:|||||:|||  ||:|||:|||||:|
Db  1533 ---AGKKYTFTMRRAGMGDG-----TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYR 1583

Qy  1629 DPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVGKTITVTCQGEA 1688
          :   |: ||||:|  |:|| |  ||  |||:|:|||||||||||||||||||||||||||
Db  1584 DAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQGEA 1643

Qy  1689 MIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
         ||||||||||||||||||||||||:|||||||||||||||||||
Db  1644 MIYDMNGRRLAAGRNTVVYTAQGGYYAVMVVVDGKSYVEKLAVK 1687
```

Db   1136 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWL------AHG 1186

Qy   475 NCCITAQFDY----VQP---CFGEVITRVKEKGAYAYIGSSPNSYWGEDYWSVGANAV 526
                :   :   :   :| | | |
Db   1187 GTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDH---------- 1232

Qy   527 FGVQPTFEGTSMGSYDATFLEDSYNTVN-------------------SIMWAGNLAATHAGN 569
           : |: | |: | | : |:/:| :|
Db   1233 YAVMISKTGTNAGDFTVVF-EETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGT 1291

Qy   570 IGNITHIGAHY--WEAYHVLGDGSVMPYRAMPKTNTYTLPASLPQNQASYISQASAGSY 627
           ||| : : | | :| ::
Db   1292 ----KIVAFRHYNCSDLNYILLDDIQFTMGGSPTFTDYT--------------- 1326

Qy   628 VAISKDGVLYGTGVANAS-----GVATVSMTKQITENGYDVVITRSNYLPVIKQIQVG-- 681
               :   : :   : | :| | ::| :|:|
Db   1327 YTVYRDGTKIKEGLTETTFEEDGVAT----------GNHEYCVEVKYTAGVSPKECVNVT 1376

Qy   682 -EPSYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPANDVRANE 740
            :| ||: |: | ||:|::: |::|
Db   1377 VDPVQNPVQNLTGSAVGQKVTLKWDAPNG---------------TPNPNPNPNP 1416

Qy   741 AKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTG----TASSNLYSANFE- 793
             |:: | | :|| |:| | | |: |:| | | | |::|: ::
Db   1417 GTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASY-INFEG 1475

Qy   794 -----YLVP------------------------------ANA--DPVV 804
                 || | || | |
Db   1476 PQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVL 1535

Qy   805 TTQNII----------VTG---QGEVIPGG------------------ 822
           ||| |: |||| |
Db   1536 TAKTVVTAPEAIRGTRVQGTMYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGK 1595

Qy   823 ----VYDYCITNEPA----------------------SGKM-WIAGDGGN------------ 846
               ||: | :   ::| |::| |
Db   1596 RADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTNVVASFSWNGM 1655

Qy   847 --QPARY--DDFT--FE------- 857
             ::|:| |::   :
Db   1656 ALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGIN 1715

Qy   858 -------------AGKKYTFTMRRAGMGDGTDMEVED---------- 881
                         | |:| | : :: ::
Db   1716 KGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYV-AFRHYNCSDLNYILLDDIQFTMGG 1774

Qy   882 -DSPASYTTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVE 940
            :::| | |:||:||||| |:| ||||||||| |||| | || |||||| | | |
Db   1775 SPTFPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVD 1834

Qy   941 GSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPGTTLSESFENGIPASWKTI 1000
           |:| |:||||| |: |||||||||||||:| | | |:|||||||||| ||:|||||
Db   1835 -PVQNPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGT-TTLSESFENGIPASWKTI 1892

Qy   1001 DADGDGHGWKPGNAPG---IAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGK 1057
           |||||| | ||| | |:|| ||| ||||| | |:||||||:||| | ||||||
Db   1893 DADGDGNNWTTTPPGGTSFAGHNSAICVSSASY-INFEGPQNPDNYLVTPELSLPGGGT 1951

Qy   1058 LTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEEIITAKG-VRSPKAIRG-RIQGT 1115
           |||||||||||||||||||:||||||||||||:||||| |:||| |||| || |:|||
Db   1952 LTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGT 2011

Qy   1116 WRQKTVDLPAGTKYVAFRHFGCTDFFWINLDEVEIKANGKRADFTETFESSTHGEAPAEW 1175
           | ||||:||||||||||||||||| |:|:|||:|||||||||||| ||||||||||||||
Db   1176 WYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIKANGKRADFTETFESSTHGEAPAEW 1235

Qy   2012 WYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIKANGKRADFTETFESSTHGEAPAEW 2071
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2072 TTIDADGDGQWLCLSSGQLDWLTAHGGSNVVASFSWNGMALNPDNYLISKDVTGATKVK 2131

Qy   1236 YYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGARFGLSTEANGAKPQSV 1295
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   2132 YYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGARFGLSTEADGAKPQSV 2191

Db   1296 WIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTFTDYTYTVYRDGTKIKEG 1355
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2192 WIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTFTDYTYTVYRDGTKIKEG 2251

Db   1356 LTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDA 1415
           ||||||||||||||||||||||||||||||:|||:|||:| |||| |||||||||||||
Qy   2252 LTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTINPTQFNPVQNLTAEQAPNSMDA 2311

Db   1416 ILKWNAPASKRAEVLNEDFENGIPASWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVS 1475
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2312 ILKWNAPASKRAEVLNEDFENGIPSSWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVS 2371

Db   1476 SASHINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 1535
           ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2372 SASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 2431

Db   1536 ANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 1595
           ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Qy   2432 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 2491

Db   1596 DVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPNGESAIETA 1655
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2492 DVVITSGNAPSYTYTIYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPNGESAIETA 2551

Db   1656 TLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYA 1715
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2552 TLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVTAQGGHYA 2611

Db   1716 VMVVVDGKSYVEKLAVK 1732
           |||||||||||||||||
Qy   2612 VMVVVDGKSYVEKLAVK 2628

RESULT 15
US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
;   APPLICANT: Travis, James
;   APPLICANT: Imamura, Takahisa
;   APPLICANT: Potempa, Jan
;   TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
;   FILE REFERENCE: 235.00160101
;   CURRENT APPLICATION NUMBER: US/09/482,500A
;   CURRENT FILING DATE: 2001-12-17
;   PRIOR APPLICATION NUMBER: US 60/115,869
;   PRIOR FILING DATE: 1999-01-13
;   NUMBER OF SEQ ID NOS: 4
;   SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
;   LENGTH: 1477
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-09-482-500A-1

Query Match   45.1%; Score 4135.5; DB 4; Length 1477;
Best Local Similarity   55.3%; Pred. No. 1.7e-275;
Matches 876; Conservative 156; Mismatches 353; Indels 199; Gaps 38;

Qy   240 TPV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTNASIKAFIHK 293
           || | |:||| :|:   : : : ::| |::| : || ::|:|| :
Db   2 TPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAIQQFVKQ 61

Qy   294 KY---NDGLAASAAPVFLALVGDTDVISGE--KGKKTKVTDLYYSAVDGDYPPEMYTF 347
           || : || | :|:| ||| || |:| : :| |: | :| :| | |
Db   62 EYEKEGND-------LTYVLLVGDHKDIPAKITPGIKSDQV---YGQIVGNDHYNEVFIG 111

Qy   348 RMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKY-GMQY 406
           |::| |||:| :::||:|:|:|||||||||:|||:| || | :|:|| :|||:| |||
Db   112 RFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQHENVIA 171

Qy   407 YYNQEHGYTDVYNYLKAPYTGCY---------SHLNTGVSFANYTAHGSETAWADPLLT 456

```
Db   172 NLLTQYGYTKIIK-------CYDPGVTPKNIIDAFNGGISLVNYTGHGSETAWGTSHFG 223
         ::|||      :|||  |||||||
Qy   457 TSQLKALTNKDKYFLAIGNCITAQFDYVQPCFGEVITRV----KEKGAYAYIGSSPNSY 512
         :: :||     ::  |:  :  |:
Db   224 TTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFARALMRAQKDGKPTGTVAIIASTINQS 283

Qy   513 WGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNI-- 570
         :
Db   284 W-------ASPMRG--------QDEMNEI-------LCEKHPNNIKR 308

Qy   571 --GNITHIGAHYYWEAYH----------VLGDGSVMPYRAMPKINTYTLPASLPQNQAS 617
                                 ::      :
Db   309 TFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQINLTDAS 368

Qy   618 YSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIK 676
         ::           |:  ::  |  |:  :  :|      :||
Db   369 VNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVVGYNKETVIK 426

Qy   677 QIQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPSAK---KAEGSREVKRIGDGLFVTIEP 732
         :  : |||||:||||||||||||||||||:|||: |      ::
Db   427 TINTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSD 486

Qy   733 AND-VRANEAKVVLAADNVWGDNTGYQFLDADHNTFGSVIPA-TGPLFTG-TASSNLYS 789
         |:  ::   ::|: :: |||:| :||:|||||: :| ||: : :  ||:|  |:::|
Db   487 APELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQYGQVIPSDTHTLWPNCSVPANLF- 545

Qy   790 ANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGGNQPA 849
         | ||| |||||||   |:| |||||| ||||  : ||  :  ||  : ::||||    |
Db   546 APFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQG---PT 601

Qy   850 RYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEE 909
          :|| :||||||| | : ::|:|||| :::| :|||||||:|||||||||||||||||||
Db   602 KEDDYVFEAGKKYHFLMKKMGSGDGTELIISEGGGSDYTYTVYRDGTKIKEGLTATTFEE 661

Qy   910 DGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNG 969
         ||||:|||||||||||||||||||||||||||||||||||||||:|:||||||||||||||
Db   662 DGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNG 721

Qy   970 TPNPNPNPNPG-TTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGCVYS 1028
         ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   722 TPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGCVYS 781

Qy  1029 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1088
         ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db   782 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 841

Qy  1089 TNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEV 1148
         |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db   842 TNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEV 901

Qy  1149 EIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVS 1208
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|:
Db   902 EIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVA 961

Qy  1209 SFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFE 1268
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   962 SFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFE 1021

Qy  1269 ETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDI 1328
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1022 ETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDI 1081

Qy  1329 QFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKC 1388
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db  1082 QFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKEC 1141

Qy  1389 VDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAE----------------- 1428
         |:|| |  |||||:||  :    | |:||| ||:||: :|
Db  1142 VNVTINPTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTEGSREVKRIGDGLFVTIEBPAN 1199

Qy  1429 -------VLNEDFENGIPASWKTIDADGDGNNWTTPPPGGSSFAGHNSAICVSSASH 1479
                 ::                                        ::
```

```
Db  1200 DVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASSNLYSA-- 1255
                                                                         :
Qy  1480 INPEG--PQNPD------NYLVT--PELSLPGGGTLTFWVCAQDANYASEHYAVYASSTG 1529
          |:  :                          :  |::|
Db  1256 -NFEYLIPANADPVVTTQNIIVTGQGEVVIPGG---VYDYCITNPEPASGKMWI-AGDGG 1310

Qy  1530 NDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-AFRHFGCTD 1588
                  :  :                                     :
Db  1311 NQPARYDDFTFE-------------------AGKKYTFTMRRAGMGD 1338

Qy  1589 FFWINLDDVVITSGNAPSYTTIYRNNTQIASGVTETTYRDPDLATGFYTYGVKVVYPNG 1648
          : ||:|||||||||||||||||||:||:||||||||||||:|||||||||||||||||||
Db  1339 G-----TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVEVKYAAG 1393

Qy  1649 ESAIETATLNITSLADVTAQKPYTLTVVGKTITVVTCQGEAMIYDMNGRRLAAGRNTVYT 1708
                          |||||||||||||||||||||||||||||||||||||||||||||||
Db  1394 VSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVVTCQGEAMIYDMNGRRLAAGRNTVYT 1453

Qy  1709 AQGGHYAVMVVVDGKSYVEKLAVK 1732
         |||| |||||||||||||||||||
Db  1454 AQGGYYAVMVVVDGKSYVEKLAVK 1477
```

Search completed: May 18, 2004, 11:36:55
Job time : 36.7501 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:32:49 ; Search time 69.1557 Seconds
(without alignments)
6969.043 Million cell updates/sec

Title: US-08-353-485-10
Perfect score: 9179
Sequence: 1 MRKLLLIAASLLGVGLYAQ.........HYAVMVVVDGKSYVEKLAVK 1732

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

SUMMARIES

Database :   Published Applications AA:*
1:    /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2:    /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3:    /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4:    /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5:    /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6:    /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7:    /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8:    /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9:    /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10:   /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11:   /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12:   /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13:   /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14:   /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15:   /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16:   /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17:   /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18:   /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	DB	Length	ID	Description
1	9167	99.9	14	1732	US-10-229-066-11	Sequence 11, Appl
2	4272	46.5	14	1706	US-10-229-066-10	Sequence 10, Appl
3	2702	29.4	15	509	US-10-387-977-101	Sequence 101, App
4	2254	24.6	15	419	US-10-174-695-5	Sequence 5, Appli
5	1776	19.3	15	419	US-10-174-695-3	Sequence 3, Appli
6	1241	13.5	15	231	US-10-174-695-6	Sequence 6, Appli
7	541	5.9	15	736	US-10-387-977-105	Sequence 105, App
8	527.5	5.7	15	491	US-10-387-977-100	Sequence 100, App
9	521.5	5.7	15	196	US-10-174-695-4	Sequence 4, Appl
10	410.5	4.5	15	507	US-10-387-977-99	Sequence 99, Appl
11	309	3.4	15	1728	US-10-282-122A-56997	Sequence 56997, A
12	303.5	3.3	12	2435	US-10-282-122A-47453	Sequence 47453, A
13	300.5	3.3	12	2468	US-10-282-122A-66335	Sequence 6635, A
14	300.5	3.3	14	2468	US-10-246-330-4	Sequence 4, Appli
15	293	3.2	12	1638	US-10-206-576-258	Sequence 258, App
16	293	3.2	12	1638	US-10-206-576-262	Sequence 262, App
17	293	3.2	12	1638	US-10-206-576-266	Sequence 266, App
18	278.5	3.0	12	6310	US-10-282-122A-67793	Sequence 67793, A
19	264	2.9	15	1283	US-10-369-493-22616	Sequence 22616, A
20	257	2.8	15	4327	US-10-369-493-10178	Sequence 10178, A
21	247.5	2.7	12	1325	US-10-282-122A-44953	Sequence 44953, A
22	247	2.7	12	3930	US-10-282-122A-46817	Sequence 46817, A
23	239	2.6	9	1741	US-09-971-536-68	Sequence 68, Appl
24	238	2.6	15	1649	US-10-369-493-18460	Sequence 18460, A
25	237.5	2.6	9	2353	US-09-797-862-33	Sequence 33, Appl
26	237.5	2.6	12	3217	US-10-282-122A-61210	Sequence 61210, A
27	232.5	2.5	12	1751	US-10-282-122A-53633	Sequence 53633, A
28	231	2.5	12	1308	US-10-282-122A-44566	Sequence 44566, A
29	229.5	2.5	12	1483	US-10-282-122A-51483	Sequence 51483, A
30	227	2.5	12	3705	US-10-282-122A-77944	Sequence 77944, A
31	226.5	2.5	12	2457	US-10-282-122A-49854	Sequence 49854, A
32	226	2.5	12	5559	US-10-282-122A-75177	Sequence 75177, A
33	225.5	2.5	12	1333	US-10-282-122A-55546	Sequence 55546, A
34	225	2.5	9	26926	US-09-759-508B-2	Sequence 2, Appli
35	224	2.5	14	1521	US-10-238-075-909	Sequence 909, App
36	223.5	2.4	12	1920	US-10-282-122A-71413	Sequence 71413, A
37	223	2.4	10	1541	US-09-839-996-3	Sequence 3, Appli
38	223	2.4	12	1541	US-10-645-655-3	Sequence 3, Appli
39	223	2.4	14	1541	US-10-080-505-3	Sequence 3, Appli
40	223	2.4	15	4283	US-10-080-334-275	Sequence 275, App
41	222.5	2.4	12	1289	US-10-282-122A-44901	Sequence 44901, A
42	221.5	2.4	12	2234	US-10-282-122A-46565	Sequence 46565, A
43	220	2.4	14	1539	US-10-085-959-253	Sequence 253, App
44	219.5	2.4	12	1953	US-10-282-122A-43730	Sequence 43730, A
45	219	2.4	12	2902	US-10-282-122A-59042	Sequence 59042, A

ALIGNMENTS

RESULT 1
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric C.
;   APPLICANT: Bhogal, Peter S.
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;   FILE REFERENCE: Reynolds
;   CURRENT APPLICATION NUMBER: US/10/229,066
;   CURRENT FILING DATE: 2002-08-28
;   PRIOR APPLICATION NUMBER: US/09/066,330
;   PRIOR FILING DATE: 1998-09-15
;   PRIOR APPLICATION NUMBER: PN 6275
;   PRIOR FILING DATE: 1995-10-30
;   PRIOR APPLICATION NUMBER: PCT/AU96/00673
;   PRIOR FILING DATE: 1996-10-30
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1732
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

Query Match   99.9%; Score 9167; DB 14; Length 1732;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1729; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRKLLLIAASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQFDASFSFNEVELTKVETK 60

Qy   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRKSFTEQVYSLNQYGSEKL 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGTFASVSIPGAFPTGEVGSPEVPAVRKLIAVPVGATPVVRKSFTEQVYSLNQYGSEKL 120

```
Qy   121  MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  MPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQYD  180

Qy   181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVVTDHGDLYNT  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVVTDHGDLYNT  240

Qy   241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  PVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA  300

Qy   301  ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI  360

Qy   361  IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNY  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  IDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNY  420

Qy   421  LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  LKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLAIGNCCITA  480

Qy   481  QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGS  540

Qy   541  YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP  600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSVMPYRAMP  600

Qy   601  KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  KTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG  660

Qy   661  NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  NYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPSAKKAEGSREVK  720

Qy   721  RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  RIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFT  780

Qy   781  GTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWI  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  GTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWI  840

Qy   841  AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  AGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKE  900

Qy   901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLFGSSVGQKV  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLFGSSVGQKV  960

Qy   961  TLKWDAPNGTPNPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY  1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  TLKWDAPNGTPNPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGY  1020

Qy  1021  NSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  NSNGCVYSESRGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS  1080

Qy  1081  TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM  1140

Qy  1141  FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTA  1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  FYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTA  1200

Qy  1201  HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNA  1260

Qy  1261  GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL  1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  GDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL  1320

Qy  1321  NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT  1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  NYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT  1380

Qy  1381  AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA  1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  AGVSPKKCVNVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRABVLNEDFENGIPA  1440

Qy  1441  SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSLP  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  SWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSSASYINFEGPQNPDNYLVTPELSLP  1500

Qy  1501  GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEVLTAKTVVTAPEAIRGTR  1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEVLTAKTVVTAPEAIRGTR  1560

Qy  1561  AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTYRNNTQIAS  1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  AQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTYRNNTQIAS  1620

Qy  1621  GVTETTYRDPDLIATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI  1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621  GVTETTYRDPDLIATGFYTYGVKVVYPNGESAIETATLNITSLADVTAQKPYTLTVVGKTI  1680

Qy  1681  TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK  1732
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681  TVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK  1732
```

```
RESULT 2
US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
;  APPLICANT: Reynolds, Eric C.
;  APPLICANT: Bhogal, Peter S.
;  APPLICANT: Slakeski, Nada
;  TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;  FILE REFERENCE: Reynolds
;  CURRENT APPLICATION NUMBER: US/10/229,066
;  CURRENT FILING DATE: 2002-08-28
;  PRIOR APPLICATION NUMBER: US/09/066,330
;  PRIOR FILING DATE: 1998-09-15
;  PRIOR APPLICATION NUMBER: PN 6275
;  PRIOR FILING DATE: 1995-10-30
;  PRIOR APPLICATION NUMBER: PCT/AU96/00673
;  PRIOR FILING DATE: 1996-10-30
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: Patentin Ver. 2.0
;  SEQ ID NO 10
;  LENGTH: 1706
;  TYPE: PRT
;  ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match         46.5%;  Score 4272;  DB 14;  Length 1706;
Best Local Similarity  50.9%;  Pred. No. 0;
Matches  932;  Conservative 204;  Mismatches 456;  Indels 240;  Gaps 47;
```

```
Qy      6  LLIAASLLGVGLYAQSAKIKLDAPTRTTCNNSFKQFDASFSNEVELIKVETKGGTFA  65
           : :||||    :  ::  :::  :|||:|: |:|:|::|: :|:|  |:|:|:|:
Db     10  IALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTEVQTPKGI--  65

Qy     66  SVSIPGAFPT------GEVGSPEVPAVRKLIAVPVGATPVRVKS--FTEQVYSLNQYG  116
           :  :||||      :  ::    |:    :   :::  | |:
Db     66  -----GQVPFTVTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSSKFIEKKNVL----  116
```

```
Qy  117 SEKLMPHQPSMSKSDDPEKVPFVYNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINP 176
            ::|:  :|:.::    |||  :   ||.|.:
Db  117 ---IAPSKGMIRNEDPKKIPYVI-GKTYSQNKFFPGEIATLDDPFILRDVRGQVVNFAP 172

Qy  177 VQYDVVANQLKVRNNIEI----EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTD 233
            :|:.:   ||| .                    :            :.:
Db  173 LQYNPVTKTLRIYTEITVAVSETSEQGKNILNKKGTFAG-----FEDTYKRMF-----MN 222

Qy  234 HGDLYNTPV------RMLVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASI 287
            ||| ||||       ||  :    :.  :    ::  :::      ::
Db  223 YEPGRYTPVEBKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAI 282

Qy  288 KAFIHKKY----NDGLAASAAPVFLALVGDTDVISGE-KGKKTKKVTDLYYSAVDGDYF 341
            :  :  :|    |||  ::..::.:            :            :
Db  283 QQFVKQEYEKEGND------LTYVLLIGDHKDIPAKITPGIKSDQV---YGQIVGNDHY 332

Qy  342 PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIK 401
            :  :             .:  :           :|  .    :   .|  |.||.||:
Db  333 NEVFIGRFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQ 392

Qy  402 Y-GMQYYYNQEHGTDVYNYLKAPYTGCY---------SHLNTGVSFANYTAHGSETAW 450
            : .:|||    |||     ||||                 :|| || ||| |||||||
Db  393 HENVIANLLTQYGYTKIIK------CYDPGVTPKNIIDAFNGGISLANYTGHGSETAW 444

Qy  451 ADPLLTTSQLKALINKDKYFLAIGNCCITAQFDVYQPCFGEVITRV----KEKGAYAYIG 506
            .  .  ::.  :|  .   :.:    :|          |  |.:    :||:|   |::|
Db  445 GTSHFGTTHVKQLINSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIA 504

Qy  507 SSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH 566
            :  .||     .|                           ||
Db  505 STINQSW-------ASPMRG------------QDEMNEI------LCEKH 529

Qy  567 AGNI----GNITHIGAHYWEAYH---------VLGDGSVMPYRAMPKINTYTLPASL 611
            ||.|    ||  :||.:||                 |||||.           |.:
Db  530 PNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQYTAPAQI 589

Qy  612 PQNQASYSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN 670
            |           .|||||:|| |.      .   ||     :        |  ||:
Db  590 NLTDASVNVSCDYNGAIATISANGKMFGSAVVE-NGTATINLI-GLTNESTLILJTVVGYN 647

Qy  671 YLPVIKQIQV-GEPSPYQPVSNLTATTQQOKVTLKWEAPSAK---KAEGSREVKRIGDGL 726
            : ||:|    |||.|||||||||||||||:|||||||||.:|   ||
Db  648 KETVIKTINTNGEPNPYQPVSNLTATTQQQKVTLKWDAPSTKINATTNTARSVDGIRELV 707

Qy  727 FVTIEPAND-VRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TA 783
            ::.:..|.| .:|:|| :|| |:||.|.:| :||.|||||: ..||||. |||:||| .:
Db  708 LISVSDAPELLRSGQAEIVLEAHDVWNDSGSYQLLLDADHDQYSQVIPSDTHTLWPNCSV 767

Qy  784 SSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGD 843
            ..||::||||| ||.|||| . .|:||.|:|.| |||.||||:||||||||||| ||||:
Db  768 PANLF-APFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQ-ANAKIWIAGQ 825

Qy  844 GGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLT 903
            |   |||||||||||||||.|.|::.|:|||| :::||||||||||||||||||||||||
Db  826 G---PTKEDDYVFEAGKKYHFLMKKMGSGDGTELITISEGGSDYTYTVYRDGTKIKEGLT 882

Qy  904 ATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLK 963
            |||||||||..||||||||||||||||||||||||||||||||||||||||:||||||||
Db  883 ATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK 942

Qy  964 WDAPNGTPNPNPNPNPG--TTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1020
            |||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db  943 WDAPNGTPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGY 1002

Qy 1021 NSNGCVYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1003 NSNGCVYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1062

Qy 1081 TGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
            |||||||||||||||||||||||||| |:|||||||||||||||||||||||||||||||
Db 1063 TGNDASNFTNALLEETITAKGVRSPEAMRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1122
```

```
Qy 1141 FYIDLDEVEIKANGKKRADFTETFBSSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1123 FYIDLDEVEIKANGKKRADFTETFBSSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTA 1182

Qy 1201 HGGSNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1260
            |||.|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1183 HGGTNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNA 1242

Qy 1261 GDFTVVFEETPNGINKGGARPGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1320
            |||||||||||||||||||||||||||||.||||||||||||||||||||||||||||||
Db 1243 GDFTVVFEETPNGINKGGARPGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 1302

Qy 1321 NYILLDDIQFTMGGSPTPTDYYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1303 NYILLDDIQFTMGGSPTPTDYYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYT 1362

Qy 1381 AGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDALLKWNAPASKRAE------- 1428
            |||||||||:|||||||||||:||.|  |.|  :.   :||:|||:
Db 1363 AGVSPKKCVNVTVNSTQFNPVKNLKAQ--PDGGDVVLKWEAPSAKKTEGSSREVKRIGDGL 1420

Qy 1429 -------VLNEDFENGIPASWKTIDADGDGNNWTTTPPPGGSSFAGHNSA 1471
                   |.:.|.||||||||||||||||.|  ||
Db 1421 FVTIEPANDVRANEAKVVLAADNVWGDNTGYQFL-LDADHNTFGSVIPATGPLFTGTASS 1479

Qy 1472 ICVSSASHINFBG--PQNPD-----NYLVT-PELSLPGGGTLTFWVCAQDANYASEHY 1521
            :         ::|  ||         ||.   |.|.|||| |||||||||||||||||
Db 1480 DLYSA---NFESLIPANADPVVTQNIIVTGQGEVVIPGG---VDYCITNPEPASGKM 1532

Qy 1522 AVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-A 1580
            |. :   |:.|.   |.|     ||                        AGKKYTFT
Db 1533 WI-AGDGGNQPARYDDFTFE--------------------AGKKYTFT 1559

Qy 1581 FRHFGCTDFFWINLDDVVITSGNAPSYTYTYRNNTQIASGVTETTYRDPDLATGFYTYG 1640
            :|:.|       ::. ::::    ||||||| |:. :.:|||||||||.  :|:|  | |
Db 1560 MRRAGMGDG----TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYC 1614

Qy 1641 VKVVVPNGESAIETATLNITSLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAA 1700
            |:|    .|    :  | : |:.||||||||||||||||:||||||||||||||||||||
Db 1615 VEVKYTAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKIITVTCQGEAMIYDMNGRRLAA 1674

Qy 1701 GRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
            |||||||||| .|||||||||||||||||.:|
Db 1675 GRNTVVYTAQGYYAVMVVVDGKSYVEKLAIK 1706
```

Query Match   29.4%; Score 2702; DB 15; Length 509;
Best Local Similarity 100.0%; Pred. No. 6.3e-199;
Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   229 DVYTDHGDLYNTPVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIK 288
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 DVYTDHGDLYNTPVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIK 60

Qy   289 AFIHKKYNDGLAASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFR 348
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AFIHKKYNDGLAASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFR 120

Qy   349 MSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYY 408
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 MSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKYGMQYYY 180

Qy   409 NQEHGYTDVYNLKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDK 468
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 NQEHGYTDVYNLKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDK 240

Qy   469 YFLAIGNCCITAQFDYYQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFG 528
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 YFLAIGNCCITAQFDYYQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFG 300

Qy   529 VQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVL 588
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 VQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVL 360

Qy   589 GDGSVMPYRAMPKTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVA 648
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GDGSVMPYRAMPKTNTYTLPASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANASGVA 420

Qy   649 TVSMTKQITENGNYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAP 708
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TVSMTKQITENGNYDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAP 480

Qy   709 SAKKAEGSREVKRIGDGLFVTIEPANDVR 737
             |||||||||||||||||||||||||||||
Db   481 SAKKAEGSREVKRIGDGLFVTIEPANDVR 509
```

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US2003023022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-5

Query Match   24.6%; Score 2254; DB 15; Length 419;
Best Local Similarity 99.8%; Pred. No. 1.4e-164;
Matches 418; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy   738 ANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLVP 797
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db     1 ANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLIP 60
```

```
Qy   798 ANADPVVTIQNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGGNQPARYDDFTFE 857
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ANADPVVTIQNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGGNQPARYDDFTFE 120

Qy   858 AGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYVYRDGTKIKEGLTATTFEEDGVAAGNH 917
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYVYRDGTKIKEGLTATTFEEDGVAAGNH 180

Qy   918 EYCEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNP 977
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 EYCEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNP 240

Qy   978 NPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGCVYSESFGLGGIG 1037
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 NPNPGTTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGCVYSESFGLGGIG 300

Qy  1038 VLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETI 1097
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 VLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETI 360

Qy  1098 TAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKR 1156
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKR 419
```

RESULT 5
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US2003023022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

Query Match   19.3%; Score 1776; DB 15; Length 419;
Best Local Similarity 79.2%; Pred. No. 9.4e-128;
Matches 336; Conservative 31; Mismatches 47; Indels 10; Gaps 6;

```
Qy   738 ANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPA-TGPLFTG-TASSNLYSANFEYL 795
             : |||:|:|| ::|| :||:|   :|  :
Db     1 SGQAEIVLEAHDVWNDSGYQILLDADHDQYGQVIFSDTHTLWPNCSVPANLF-APFEYT 59

Qy   796 VPANADPVVTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFT 855
             || ::|:|   :| :  ::
Db    60 VPENADPSCSPTNMIMDGTASVNIPAGTYDFPALAAFQ-ANAKIWIAGQG---PTKEDDYV 115

Qy   856 FEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTYVYRDGTKIKEGLTATTFEEDGVAAG 915
             ||||||| | :: ::||||| :|    ||  :|||||||||||||||||||||||||||||
Db   116 FEAGKKYHFLMKKMGSSGDGTELTISEGGGSDYTYTYVYRDGTKIKEGLTATTFEEDGVAAG 175

Qy   916 NHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNP 975
             |||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db   176 NHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNP 235

Qy   976 NPNPNPG---TTLSESFENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGCVYSESFG 1032
             |||||
Db   976
```

```
Db   236 NPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFG 295

Qy  1033 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1092
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   296 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 355

Qy  1093 LEETITAKGVRSPKAIRGRIQGTWROKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVBIKA 1152
         |||||||||:||:|||:|||||||||||||||||||||||||||||||||||||||||||
Db   356 LEETITAKGVRSPEAIRGRIQGTWROKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVBIKA 415

Qy  1153 NGKR 1156
         ||||
Db   416 NGKR 419
```

RESULT 6
US-10-174-695-6
; Sequence 6, Application US/10174695
; Publication No. US20030233222A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NOS: 8
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-6

```
Query Match          13.5%;   Score 1241; DB 15;   Length 231;
Best Local Similarity 99.6%;   Pred. No. 5.5e-87;
Matches 230;  Conservative 1;   Mismatches 0;  Indels 0;   Gaps 0;

Qy   759 FLLDADHNTPGSVIPATGPLFTGTASSNLYSANFEYLVPANADPVVTTQNIIVTGQGEVV 818
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 FLLDADHNTPGSVIPATGPLFTGTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVV 60

Qy   819 IPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDME 878
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 IPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDME 120

Qy   879 VEDDSPASYTYTVRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVT 938
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 VEDDSPASYTYTVRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVT 180

Qy   939 VEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLSESF 989
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 VEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLSESF 231
```

RESULT 7
US-10-387-977-105
; Sequence 105, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200000301

; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 105
; LENGTH: 736
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-105

```
Query Match           5.9%;   Score 541; DB 15;   Length 736;
Best Local Similarity 25.1%;  Pred. No. 3.2e-32;
Matches 200;  Conservative 130;  Mismatches 328;  Indels 140;  Gaps 29;

Qy     3 KLLLLIA-ASLLGVGLYAQSAKIKLDAPTTRTTCTNNSFKQPDASFSFNEVELTKVETKG 61
         ::::|:: :||||: :|||  |  |||:
Db     7 RIVSIVAFSSLLGGMAFAQPAR-RGRNPQVRLLSARQSMSK--VQFRMDNLQFTGVQTSK 63

Qy    62 GTFASVSIPGAPPTGEVGSPEVPAVRKLIAVPVGATPVVRVKS--FTEQVYSLNQYGSEK 119
         |:||::::: ||:
Db    64 GVAQVPFFTEGVNISEKGTPILPILSRSLAVSETRAMKVEVVSSKFIEKKDVL------- 116

Qy   120 LMPHQPSMSKSDDPEKVPFVNAAAYARKGFVGQELTQVEMLGTMRGVRIAALTINPVQY 179
         :: |:  ::
Db   117 IAPSKGVISRAENPDQIPYVVGQSYNEDKFFPGEIATLSDPF-ILRDVRGQVVNFAPLQY 175

Qy   180 DVVANQLKVRNNIEIEVSFQGADEVATQRLVDASFSPYFETAYKQLFNRDVVTDHGDLYN 239
         : :
Db   176 NPVTKTLRIYTBIVAVSETAEAGQNTISLVKNSTFTGFEDIYKSVFMNYEATRY------ 230

Qy   240 TPV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHK 293
         :|: ::  ::
Db   231 TPVEEKENGRMIVIVPKKYEEDIEDFVDWKNQRGLRTEVKVAEDIASPVTANAIQQFVKQ 290

Qy   294 KY----NDGLAASAAPVFLAIVGDTDVISGE--KGKKTKKVTDLLYSAVDGDYFPEMYTF 347
         ||   :::|: :::  |:|
Db   291 EYEKEGND------LTYVLLVGDHKDIPAKITPGIKSDQV--YGQIVGNDHYNEVFIG 340

Qy   348 RMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKY-GMQY 406
         |:|| ::  : :  :: :
Db   341 RFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQHENIIA 400

Qy   407 YYNQEHGYTDVYNYLKAPYTGCY----------SHLNTGVSFANYTAHGSETAWADPLLT 456
         :: : : :
Db   401 NLLTQYGYTKIIK----------CYDPGVTPKNIIDAFNGGISLANYTGHGSETAWGTSHFG 452

Qy   457 TSQLKALTNKDKYFLAIGNCCITAQPDYVQPCFGEVITRV----KEKGAYAYIGGSPNSY 512
         :|:::: ::  :: :
Db   453 TTHVKQLTNSNQLPFIFDVACVNGDFLYNVPCFAEALMRAQQKDGKPTGTVAIIASTINQS 512

Qy   513 WGEDYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNI-- 570
         |  :: ::  : :
Db   513 W------ASPMRG------------QDEMNEI------------LCEKHPNNIKR 537

Qy   571 --GNITHIGAHYYWEAYH------------VLGDGSVMPYRAMPKTNTYTLPASLPQNQAS 617
         :| :|
Db   538 TFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPANISASAQT 597

Qy   618 YSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIK 676
         :: |
Db   598 FEVACDYNGAIATLSDDGDMVGTAIVK-DGKAIIKLNESIADETNLTLTVVGYNKVTVIK 656

Qy   677 QIQVGEPSPYQPVSN---LTATTQSQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPA 733
         :|: : :|: |
Db   657 DVKV-EGTSIADVANDKPYTVAVSGKTITV--ESPAA----------GL-TIFDM 697

Qy   734 NDVRANEAK--VVLAADN 749
```

Db   698 NGRRVATAKNRMVFEAQN 715

RESULT 8
US-10-387-977-100
; Sequence 100, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928000301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 100
;   LENGTH: 491
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-100

Query Match   5.7%;   Score 527.5; DB 15;   Length 491;
Best Local Similarity   28.8%;   Pred. No. 1.8e-31;
Matches   153; Conservative   82; Mismatches 190; Indels 107; Gaps   19;

Qy   240 TPV------RMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHK 293
              ||: |:|:::| ::| :| :| :|:|||:|:| :|:|:|
Db   2 TPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAIQQFVKQ 61

Qy   294 KY-----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYYSAVDGYFPEMYTF 347
             :| || : ||| :| :|| :||| ||| | ||
Db   62 EYEKEGND--------LTYVLLIGDHKDIPAKITPGIKSDQV--YGQIVGNDHYNEVFIG 111

Qy   348 RMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKY-GMQY 406
              ||:|| : |::| ||| :| :: ||:: |::| :::
Db   112 RFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGESDIQHENVIA 171

Qy   407 YYNQEHGYTDVVNYLKAPYTGCY----------SHLNTGVSFANYTAHGSETAWADPLLT 456
               ::|||: | | ||| ||:||| :|
Db   172 NLLTQYGYTKIIK-------CYDPGVTPKNIIDAFNGGISLANYTGHGSETAWGTSHFG 223

Qy   457 TSQLKALTNKDKYFLAIGNCCITAQFDVVQPCFGEVITRVKEKG--AYAYIGSSPNSYW 513
               |||| |::||:|||::| : |: ||: :::| |||::|:: ::|
Db   224 TTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGVAIIASTINQSW 283

Qy   514 GEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNI---- 570
              || ||: | : || :   : ||
Db   284 ------ASPMRG---------QDEMNEI-------------LCEKHPNNIKRT 308

Qy   571 -GNITHIGAHYYWEAYH----------VLGDGSVMPYRAMPKTNTYTLPASLPQNQASY 618
                 :   : || |||| ::| ::| :::| |::|| :|::
Db   309 FGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQINLTDASV 368

Qy   619 SIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIKQ 677
               ::| |: | | | ||:| | : | :|:||::|:| ::|:
Db   369 NVSCDYNGAIATISANGKMFGSAVVE-NGTATINLT-GLTNESTLTLTVGYNKETVIKT 426

Qy   678 IQV-GEPSPYQPVSNLTATTQGQKVTLKWEAPS------------AKKAEGSREV 719
               |::||||||||||||||||||||||:   : :|
Db   427 INTNGEPNPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIREL 478

RESULT 9
US-10-174-695-4
; Sequence 4, Application US/1017495
; Publication No. US20030232022A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: Slakeski, Nada
;   APPLICANT: Chen, Chao Guang
;   APPLICANT: Barr, Ian George
;   TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
;   FILE REFERENCE: 52928000700
;   CURRENT APPLICATION NUMBER: US/10/174,695
;   CURRENT FILING DATE: 2002-06-18
;   PRIOR APPLICATION NUMBER: PCT/AU00/01588
;   PRIOR FILING DATE: 2000-12-21
;   PRIOR APPLICATION NUMBER: AU PQ 4859
;   PRIOR FILING DATE: 1999-12-24
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 196
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-174-695-4

Query Match   5.7%;   Score 521.5; DB 15;   Length 196;
Best Local Similarity   92.2%;   Pred. No. 1.1e-31;
Matches   94; Conservative   3; Mismatches   2; Indels   3; Gaps   1;

Qy   932 KVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPNPG--TTLSES 988
              ||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1 KVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTTLSES 60

Qy   989 FENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGCVYSES 1030
              |||||||||||||||||||||||||||||||||||| ::
Db   61 FENGIPASWKTIDADGDHGWKPGNAPGIAGYNSNGCVYLDN 102

RESULT 10
US-10-387-977-99
; Sequence 99, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928000301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 99
;   LENGTH: 507
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-99

Query Match   4.5%;   Score 410.5; DB 15;   Length 507;
Best Local Similarity   25.4%;   Pred. No. 1.9e-22;
Matches   142; Conservative   86; Mismatches 209; Indels 121; Gaps   22;

Qy   240 TPV------RMLVVAGAKFKEALKPWLITWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHK 293
              ||: |:|:::| ::| :| :| :|:|||:|:| :|:|:|
Db   2 TPVEEKENGRMIVIVPKKYEEDIEDFVDWKNQRGLRTEVKVAEDIASPVTANAIQQFVKQ 61

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56997
; LENGTH: 1728
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-10-282-122A-56997

Query Match            3.4%;   Score 309;  DB 12;  Length 1728;
Best Local Similarity 21.3%;   Pred. No. 1e-13;
Matches  375;  Conservative 163;  Mismatches 617;  Indels 604;  Gaps  89;

Qy  132 DPEKVPF--VYNAAAYARKGFVGQELTQV------EMLGTMRGVRIAALTINPVQYDVV 182
Db  137 DLSKVTFLTAVLNAA---------NDLTNVITQITSGALGNLTGVDIDLTEVN------- 180

Qy  183 ANQLKVRNNIEI--EVSFQGADEVATQRLYDASFSPYFETAYKQLFNRDVYTDHGDLYNT 240
Db  181 -RQLELVNNIENLGAASFTAPETLAADGSY--ISAPISDGLGLVLAQNVSNILQDL-NA 235

Qy  241 PVRMLVVAGAKFKEALKPWLTMKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYNDGLA 300
Db  236 AVQALEAKGTSIPSNL------------VAAAINA--ALILPVKGTVNVA 270

Qy  301 ASAAPVFLALVGDTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360
Db  271 VSGALPLLAVGG------SG------VNELVDASLLG--TTTVLPTTVSTPQNLSNN 314

Qy  361 IDKVLMYEKATMPDKSYLEKVLLIAGADYSWN------SQVGQPTIKYGMQYYYNQEH 412
Db  315 LDARFV--GTVVQTDLLD-VNLLATADGVSNIYFAAGTTSEVTAPTI------ 358

Qy  413 GYTDVYNYLKAPYTGCYSHLNTGVSFANYTAHGSETAWADPLLTTSQLKALTNKDKYFLA 472
Db  359 ------------TGV--TGNSTAGYEVKGT---ADANATVE---IRNAGGTVIG 392

Qy  473 IGNCCITAQFDYVQPC---FGEVITRVKEKGAYAYIGSSPNSYWGEDYYWSVGANAVFG 528
Db  393 TGTADGTGAFTVTVPAGEAGANETLTAVAKNASGT--ESTPTTFQTPADEATVTAPTITG 450

Qy  529 VQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVL 588
Db  451 V----TGNSTAGYEVKGTADANATVE------IRNAGGT------VI 481

Qy  589 GDGSVMPYRAMPKTNTYTLPASLPQNQASYSIQASAG-SYVAISKDGVLYGTGVANASGV 647
Db  482 GGTADGTGAF----TVTVPAG------EAGANETLTAVAK----NASGT 517

Qy  648 ATVSMTKQITENGNDVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEA 707
Db  518 ESTPTTFQ------TPADEATVTAPTITGVTGNSTAGYEV 551

Qy  708 PSAKKAEGSREVKRI------GDGLFVTIEPANDVRANEAKVVLAADNVWGDNTG 756
Db  552 KGTADANATVEIRNAGGTVIGTGTADGTGAFTVTVPAGEAGANETLTAVAKNASGTESTP 611

Qy  757 YQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLVPANADPVVTTQN---IIVT 812
Db  612 TTFQTPADEAT------VTAPTITGVTGNS--TAGYEVKGTADANATVEIRNAGGAVIGT 663

Qy  813 GQGEVVIPGGVYDYCITNPEPASGKMWIAGDGG----845
Db  664 GTAD----GTGAFTVTIP----AGEAGANETLTAVAKNASGTESTPTTFQTPAD 709

Qy  846 -NQPAR--YDDFTFEAGKKYTFMRRAGMGD-GTDMEVEDDSPASYTVTYVRDGTKIKE 900
```

```
Qy  294 KY----NDGLAASAAPVFLALVGDTDVISGE--KGKKTKKVTDLYYSAVDGDYFPEMYTF 347
Db   62 EYEKEGND------LTYVLLVGDHKDIPAKITPGIKSDQV---YGQIVGNDHYNEVFIG 111

Qy  348 RMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGADYSWNSQVGQPTIKY-GMQY 406
Db  112 RFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAEGGPSADNGRSDIQHENIIA 171

Qy  407 YYNQEHGYTDVYNYLKAPYTGCY--------SHLNTGVSFANYTAHGSETAWADPLLT 456
Db  172 NLLTQYGYTKLIK------CYDPGVTPKNIIDAFNGGISLANYTGHGSETAWGTSHFG 223

Qy  457 TSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV---KEKGAYAYIGSSPNSY 512
Db  224 TTHVKQLTNSNQLPFIFDVACVNGDFLYNVPCFAEALMRAQKDGKPTGTVAIIASTINQS 283

Qy  513 WGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNI-- 570
Db  284 W------ASPMRG--------------QDEMNEI-------LCEKGHPNNIKR 308

Qy  571 --GNITHIGAHYWEAYH----------VLGDGSVMPYRAMPKTNTYTLPASLPQNQAS 617
Db  309 TFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVIAPANISASAQT 368

Qy  618 YSIQAS-AGSYVAISKDGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIK 676
Db  369 FEVACDYNGAIATLSDDGDMVGTAIVK-DGKAIIKLNESIADETNLTLTVVGYNKVTVIK 427

Qy  677 QIQVGEPSPYQPVSN--LTATTQGQKVTLKWEAPSAKKAEGSREVKRIGDGLFVTIEPA 733
Db  428 DVKV-EGTSIADVANDKPYTVAVSGKTITV--ESPAA------GL-TIFDM 468

Qy  734 NDVRANEAK--VVLAADN 749
Db  469 NGRRVATAKNRMVFEAQN 486
```

```
RESULT 11
US-10-282-122A-56997
; Sequence 56997, Application US/10282122A
; Publication No. US2004029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

```
;   APPLICANT:  Zamudio, Carlos
;   APPLICANT:  Malone, Cheryl
;   APPLICANT:  Haselbeck, Robert
;   APPLICANT:  Ohlsen, Kari
;   APPLICANT:  Zyskind, Judith
;   APPLICANT:  Wall, Daniel
;   APPLICANT:  Trawick, John
;   APPLICANT:  Carr, Grant
;   APPLICANT:  Yamamoto, Robert
;   APPLICANT:  Forsyth, R.
;   APPLICANT:  Xu, H.
;   TITLE OF INVENTION:  Identification of Essential Genes in Microorganisms
;   FILE REFERENCE:  ELITRA.034A
;   CURRENT APPLICATION NUMBER:  US/10/282,122A
;   CURRENT FILING DATE:  2003-02-20
;   PRIOR APPLICATION NUMBER:  60/191,078
;   PRIOR FILING DATE:  2000-03-21
;   PRIOR APPLICATION NUMBER:  60/206,848
;   PRIOR FILING DATE:  2000-05-23
;   PRIOR APPLICATION NUMBER:  60/207,727
;   PRIOR FILING DATE:  2000-05-26
;   PRIOR APPLICATION NUMBER:  60/230,335
;   PRIOR FILING DATE:  2000-09-06
;   PRIOR APPLICATION NUMBER:  60/230,347
;   PRIOR FILING DATE:  2000-09-09
;   PRIOR APPLICATION NUMBER:  60/242,578
;   PRIOR FILING DATE:  2000-10-23
;   PRIOR APPLICATION NUMBER:  60/253,625
;   PRIOR FILING DATE:  2000-11-27
;   PRIOR APPLICATION NUMBER:  60/257,931
;   PRIOR FILING DATE:  2000-12-22
;   PRIOR APPLICATION NUMBER:  60/267,636
;   PRIOR FILING DATE:  2001-02-09
;   PRIOR APPLICATION NUMBER:  60/269,308
;   PRIOR FILING DATE:  2001-02-16
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS:  78614
;   SOFTWARE:  PatentIn version 3.1
;
; SEQ ID NO 47453
;   LENGTH: 2435
;   TYPE: PRT
;   ORGANISM: Burkholderia cepacia
US-10-282-122A-47453

Query Match             3.3%;  Score 303.5;  DB 12;  Length 2435;
Best Local Similarity  21.5%;  Pred. No. 4.8e-13;
Matches  309;  Conservative 121;  Mismatches 538;  Indels 467;  Gaps 68;
```

```
Qy      531 PIFEGTS-----MGSYDATFLEDSYNTVNSIMW----------AGNLAATHAGNIGNIT 574
                :                 : ::             :::          ||||:::
Db      244 PTLGGTAEAGSRVSVYDGTTLLGTTLGTTTADSSGKWTFPTTGLGEGAHSITVTATDAAGNVS 303

Qy      575 HIGAHYWEAYHVLGDGSVMPYRAMPKTNTYTLPASLPQNQA----SYSIQASAGSVVAI 630
                                                   ::
Db      304 ------------------------------TPSAAFELTIDTTAPA-LPTVNATDGTSLSGTAEAGATVNI 343

Qy      631 --SKDGVLYGTGVANASGVATVSMTKQITENGYADVVITRSNYLPVIKQIQVGEPSYQP 688
              :                        :|                    ::
Db      344 DTNGDGTPDATVTADPSGVWT---------------------------YTP-------STPLPIGT 375

Qy      689 VSNLTAT----------TOGQKVTLKWEAPSAK----------KABGSREVKRIGDG-----L 726
              :|||                             :|||
Db      376 VIGVTATDAAGNTGPSASVTVTGDTTAPGAPVIGTVTDDAGSVVGAIASGGSTDDATPTL 435

Qy      727 FVTIEPANDVRANEAKVVL----------------AADNVWGDN 754
                                                ||||
Db      436 SGTAEAGSTVSVYDGTTLLGTTLGTTTADPSGNWTPTPTTCLEEGAHSLTVTATDTAGNVSVPS 495

Qy      755 TGYQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLVPANADPVVTTQNIIVTGQ 814
              ||:
Db      496 TAFDLTIDTTAPSIPTVNATDGTSLSGTAEAG--------------ATINIDTNGD 537

Qy      815 GB-----VVIPGGVYDYCITNPEPA------------------SGKMWIAGDGGNQPARYD 852
```

(left column Db/Qy blocks, page 8 continued)

```
Db      710 PNTPVATPIVETVTGSTTKGYEV----KGTAEVGTTIEVRDAAG-------TVLGTATTGTD 760
Qy      901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVC-----KDVTVEGSNEFAP-VQNLTGS 954
Db      761 GKYTVTL-DSGIATANQTLSVVAKNASGTESQPATATTPADVT------APTVDNITGN 812
Qy      955 S-VGQKVTLKWDAPNGTPNPN-------------------PNPNPNPGTTL 985
Db      813 SGSGYEIT-------GTADPNTTIEVRDPSGAVIGTGTSDANGDFTVTLPTGTTNPGDTL 865
Qy      986 SESFENGIPASWKTIDADGDGHGWKPG---------NAPGIAGYNSNGCVYSESFGLGG 1035
Db      866 T---------------VIGKDNAGNESQPTEVLVPADATVTAPTVTGVTGNSVA-------- 904
Qy     1036 IGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFT------ 1089
Db      905 -----------GYQVTGTAD-PNA---TIEIRDADGN------VIATGTADGTGSFAVNLPAG 946
Qy     1090 NALLEEITTAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLLDEV- 1148
Db      947 TANANETLTA-------------------LAKDPAG-------NTSTPTTFQTPADEVV 979
Qy     1149 -----EIKANGKRA-DPTETEFSSTHGEAPAEWTTI----DADGDGQGWLCLSSGQLDW 1197
Db      980 APPSVDKVTGNTTQGYQVTGTAELGTTIEVRATDGTVLGTTATTGPTGQYTVTLIASGK-- 1036
Qy     1198 LTAHGGSNVVSSFSWNGMALNPDNYLIISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTG 1257
Db     1037 ATAKQTVNVVAK-NDTGLESQPTTAMTPADVTTPT-----------------IG 1072
Qy     1258 TNAGDFTVVFEET----PNGI----NKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKY 1309
Db     1073 DITGDSTGYEITGTADPNTTIEVRNPDGTIIGTTTTDDQG-----NFTVDLPAGAA----- 1124
Qy     1310 VAFRHYNCSDLNVILLDDIQFTMGGSPTPIDYTY----TVRDGTIKEGLTETTFEEDG 1365
Db     1125 ------NPGDTLTVGKDGD---GNESQPTEVTVPEDATVAAPTVTVTGTTAATGYQVTG 1175
Qy     1366 VATGNHEYCVEVKYTAG-VSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILK------ 1418
Db     1176 TAEPN--VTIEIHNEAGLVIATGTTDGAGAFTITLPTGTATANEA--LTAIAKDAAGKE 1230
Qy     1419 -----WNAPASKRAEVLNEDFE---------NG-------IPASWKTIDADG--------- 1449
Db     1231 SNPTAFKTPADPDAPVATPTVDKITGSTTNGYQVVGAAEVGTTVEVRDADGTVLGMATTG 1290
Qy     1450 -DGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTP-ELSLPGGGTLTF 1507
Db     1291 TDG-KYTVTLEPGKAS---ANETITTVVAK---NATGKESQPATATTPVDLATPTIDSIT- 1342
Qy     1508 WVCAQDANYASEHY------------------AVYASSTGNDASNFANALLEEVLT-AK 1547
Db     1343 -----GNSSKGYEITGTAEPKTTIDVRDADGTIIAATTANETGQYTVTLPAGVVTPGE 1395
Qy     1548 TVVTAPEAIRGTRAQGTWYQKTVQLPAGTKVVAFRHFGCTDFFWINLDDVVITSGNAPSY 1607
Db     1396 TITIISKDGAGNESQ---PATAVIPA--------------------DVVLA--AP-- 1425
Qy     1608 TYTIYRNNTQIASGVTVETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNITSLADVTA 1667
Db     1426 --TITKVEGNKANGYIVTGTADPNVTVQFYNSSEQLLASGNTTTGGTFSVHIA--AGLAT 1481
Qy     1668 QKPYTLTVGKTITVTCQG 1686
Db     1482 EKE-TLT----ALTTDTQG 1495

RESULT 12
US-10-282-122A-47453
; Sequence 47453, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, Liangsu
```

US-10-282-122A-66335
; Sequence 66335, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, Liangsu
;   APPLICANT: Zamudio, Carlos
;   APPLICANT: Malone, Cheryl
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Kari
;   APPLICANT: Zyskind, Judith
;   APPLICANT: Wall, Daniel
;   APPLICANT: Trawick, John
;   APPLICANT: Carr, Grant
;   APPLICANT: Yanamoto, Robert
;   APPLICANT: Forsyth, R.
;   APPLICANT: Xu, H.
;   TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;   FILE REFERENCE: ELITRA.034A
;   CURRENT APPLICATION NUMBER: US/10/282,122A
;   CURRENT FILING DATE: 2003-02-20
;   PRIOR APPLICATION NUMBER: 60/191,078
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: 60/206,848
;   PRIOR FILING DATE: 2000-05-23
;   PRIOR APPLICATION NUMBER: 60/207,727
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: 60/230,335
;   PRIOR FILING DATE: 2000-09-06
;   PRIOR APPLICATION NUMBER: 60/230,347
;   PRIOR FILING DATE: 2000-09-09
;   PRIOR APPLICATION NUMBER: 60/242,578
;   PRIOR FILING DATE: 2000-10-23
;   PRIOR APPLICATION NUMBER: 60/253,625
;   PRIOR FILING DATE: 2000-11-27
;   PRIOR APPLICATION NUMBER: 60/257,931
;   PRIOR FILING DATE: 2000-12-22
;   PRIOR APPLICATION NUMBER: 60/267,636
;   PRIOR FILING DATE: 2001-02-09
;   PRIOR APPLICATION NUMBER: 60/269,308
;   PRIOR FILING DATE: 2001-02-16
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 78614
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66335
;   LENGTH: 2468
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66335

Query Match   3.3%;   Score 300.5; DB 12; Length 2468;
Best Local Similarity 20.5%;   Pred. No. 8.4e-13;
Matches 305;   Conservative 140;   Mismatches 505;   Indels 535;   Gaps 74;

Qy   534 EGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGN--------------IGNITHIGAH 579
                     ::   :::   : :
Db   99 DGPLMAGY---IPQESLAGFESLTGAGVLGGMSAGTALLVGAAAIGAGVAISNSSGGGGG 155

Qy   580 YYWEAYHVLGDGSVMPYRAMPKTNYTLPASLPQNQASYSIQASAGSYVAI--SKDGVLY 637
                                                     :::   :
Db   156 ----------GGSSVPDTTPPKAAS--GLKIAPDGSSISGQAEAGASVGIDTNGDGKPD 203

Qy   638 GTGVANASGVATVSMTKQITENGNYDVVIT--RSNYLPVIKQIQVGEPSPYQPVSNLTAT 695
                :::   :
Db   204 LTVIADANGNFTAPLNPPLLNGQTVTVVVTDDVGSVGAI-ISG   262

Qy   696 TQGQKVTLKWEAPSAKKAEGSREVKRIGDG---------LFVTIEPANDVRA-------- 744
                     :::   :
Db   263 PDGSSVTGKAEPGSTVGVDTD------GDGQPDTTVVVGPGGSFEVPLNPPLTNGETVTV 316

Qy   745 LAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLVPANADPVV 804
                                 :: : :   :   :
Db   317 IVTDPAGNNST--PVTVEAPDTT-APAPATDVQVAPDGSS------------VTGNAEPGA 362

RESULT 13

```
Db   1285  GNWSFTPGTPLANGTVNAVAQDPAGNTG------------||||  PQSSTTVDAVAPNTPVVNPSN  1334
                   :||:|:|:.  :||
Qy   1661  -SLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNT  1704
                   :|:.||:.  ||
Db   1335  GNLLNGTAEP-----------GSTVTLIT-----------DGNGNPI--GQTT  1362
```

RESULT 14
US-10-246-330-4
; Sequence 4, Application US/10246330
; Publication No. US2003016603OA1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Fah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; TITLE OF INVENTION: BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE
; FILE REFERENCE: 14537-002001
; CURRENT APPLICATION NUMBER: US/10/246,330
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/323,241
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 2468
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-10-246-330-4

Query Match   3.3%;   Score 300.5;   DB 14;   Length 2468;
Best Local Similarity 20.5%;   Pred. No. 8.4e-13;
Matches 305;   Conservative 140;   Mismatches 505;   Indels 535;   Gaps 74;

```
Qy    534  EGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGN--------------IGNITHIGAH  579
                                                                 :
Db     99  DGPLMAGY---IPQESLAGFESLTGAGVLGGMSAGTALLVGAAAIGAGVAISNSSGGGGG  155
                  :|::
Qy    580  YYWEAYHVLGDGSVMPYRAMPKTNTYTLPASLPQNQASYSIQASAGSYVAI--SKDGVLY  637
Db    156  --------GGSSVPPDTTPPKAAS--GLKIAPDGSSISGQAEAGASVGIDTNGDGKPD    203
                  ::|
Qy    638  GTGVANASGVATVSMTKQITENGNYDVVIT--RSNYLPVIKQIQVGEPSPYQPVSNLTAT  695
                  :|:|:.                   :|
Db    204  LTVIADANGNTAPLNPPLTNGQTVTVVVTDPAGNASPP--AQVTAPDTTAPAPATDVQVA  262
Qy    696  TQGQKVTLKWEAPSAKKAEGSREVKRIGDG--LFVTIEPANDVRA---------NEAKVV  744
                  :|                           :|
Db    263  PDGSSVTGKAEPGSTVGVDTD----------GDGQPDTTVVVGPGGSFEVPLNPPLTNGETVTV  316
Qy    745  LAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLVPANADPVV  804
                  ||:|
Db    317  IVTDPAGNNST--PVTVEAPDTT--APAPATDVQVAPDGSS--------------VTGNAEPGA  362
Qy    805  TTQNIIVTGQGE-----VVIPGGVYDYCITNPEPASGK---MWIAGDGGNQ---PARYD  852
                  :||:.|:.   :||
Db    363  TV-GVDTDGDGQPDTTVVVGPGGSFEVPL-NPPLTNGETVTVIVTDPAGNSTPVTAEAP  420
Qy    853  DF------------------TFEAGKKYTFTMRRAGMGDGTDM-EVEDDSPASYTYTYRDG  895
                  ||
Db    421  DFPDAPQVNASNGSVLSGTAEAGVTIVIT-----DGNGNPIGQTSADANGNWSFT----PG  472
Qy    896  TKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPV-------Q  949
                  :|:|
Db    473  SQLPDGTVVNVVARD--AAGNS--------------SP-ATSITVDGVAPNAPVVEPSNGS  516
Qy    950  NLTGSS-VGQKVTL-----------------KWDAPNGTPNPN-----------  974
                  |:|::.
Db    517  ELSGTAEPGSSVTLTDGNGNPIGQTADANGNWSFTPSTPLPDGTVVNVVARDAAGNSSP  576
Qy    975  ----------------PNPNPNPGTTLSESFE------------NGIPASWKTIDADGDGHGW--  1009
                  :|:||
Db    577  PASVTVDAVAPATPTVDPSNGTTLSGTAEPGSSVTLTDGNGNPIG--QVTADGSGN-WTF  633
```

(left column continued, matching Db/Qy pairs, rotated sequence data)

```
Qy    805  TTQNIIVTGQGE-----VVIPGGVYDYCITNPEPASGK--MWIAGDGGNQ----PARYD  852
Db    363  TV-GVDTDGDGQPDTTVVVGPGGSFEVPL-NPPLTNGETVTVIVTDPAGNSTPVTAEAP  420
Qy    853  DF----------------TFEAGKKYTFTMRRAGMGDGTDM-EVEDDSPASYTYTYRDG  895
Db    421  DFPDAPQVNASNGSVLSGTAEAGVTIVIT----DGNGNPIGQTADANGNWSFT----PG  472
Qy    896  TKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPV-----Q  949
Db    473  SQLPDGTVVNVVARD------------AAGNS----SP-ATSITVDGVAPNAPVVEPSNGS  516
Qy    950  NLTGSS-VGQKVTL----------------KWDAPNGTPNPN-------  974
Db    517  ELSGTAEPGSSVTLTDGNGNPIGQTADANGNWSFTPSTPLPDGTVVNVVARDAAGNSSP  576
Qy    975  ---------------PNENPNPGTTLSESFE-------NGIPASWKTIDADGDGHGW--  1009
Db    577  PASVTVDAVAPATPTVDPSNGTTLSGTAEPGSSVTLTDGNGNPIG--QVTADGSGN-WTF  633
Qy   1010  ---------------KPGNAPGIAGYNSNGCVYS  1028
Db    634  TPSTPLPNGTVNATATDPSGNASSPASVTVDAVAPATPVVNPSNGTTLSGTAEPGATVT  693
Qy   1029  ESFGLGG-IGVLTPD---NYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGND  1084
Db    694  LTDGNGNPIGQVTADGSGNWSFTPTTPLPNG----------TVVNATATD  733
Qy   1085  ASNFTNALLELTI---------TAKGVRSPKAIRGRIQGTWR--QKTVDLPA  1125
Db    734  ASGNTSAGSSVTVDSVAPATPVINPSNGTTLSGTAEPGSSVTLTDGNGNPIGQVTAD---  790
Qy   1126  GTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES-STHGEAPAEWTTIDADGDG  1184
Db    791  GSGNWSFT--PSTPL--------------ADGTVVNATATDPAGNTSGQGS---TTVDGVAPT  834
Qy   1185  QGWLCLSSGQLDWLTAHGGSNVV-------------SSFSW-------NGMALNPD  1220
Db    835  TPTVNISNGSLSGTAEPGSVTLTDGNGNPIAEVTADGSGNWTYTPSTPIANGTVVN--  892
Qy   1221  NYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGAR  1280
Db    893  --VVAQDAAGNSSPGASVTVDSQAPA--APVVNPSNGT-------TLSGTAEPGAT  937
Qy   1281  FGLSTEANG--------AKPQSVWIERTVDLPAGTKVVAFRHYNCSDLNYILLDDIQFTMG  1333
Db    938  VTL--TDGNGNPIGQVTADGSGNW-------SFTPGTPLA-------NGTVVNATASDPTGNTSA  986
Qy   1334  GSPTPPDYTYTVYY------RDGTKI---KEGLTETTFEEDG------VATGNHEYCVE  1376
Db    987  PASTTVDSVAPAAPVVNPSNGAEISGTAEPGATVTLTDGSGNWSFTPS  1046
Qy   1377  VKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFEN  1436
Db   1047  IVTDPAGNNST--PVTVEAPDTT---APAPATDVQVAPDGSS  1093
Qy   1437  GIPASWKT-IDADGDGN--------NWTTTP------APATPTVNLSNGSSLS  1458
Db   1094  GTAEPGSTVILTDGNGNPIAEVTADGSGNWTYTPSTPIANGTVVNVVAQDASGNSSPPAT  1153
Qy   1459  DF----------PP----------GGSSFAGHNSAICVSSASHINFEGPQNPDNYLVTPELSL  1499
Db   1154  VTVDSSAPPAPVINPSNGVVISGTAEGATVTLTDAGGNPIGQVTADGSGNWSFTPGTPL  1213
Qy   1500  PGGGTLTFWCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVTAPEAI----  1556
Db   1214  ANGTVIV-----ATATDPTGNTGPQAA--------TTVDAVAPAPVIDP  1250
Qy   1557  -RGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYI---  1611
Db   1251  SNGTTISGT------ABAGAK---------------VILTDGNGNPIGETTADGS  1284
Qy   1612  ----YRNNTQIASG-VTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNIT---  1660
Db    577  PASVTVDAVAPATPTVDPSNGTTLSGTAEPGSSVTLTDGNGNPIG--QVTADGSGN-WTF  633
```

```
                    CITY: Rockville
                    STATE: Maryland
                    COUNTRY: USA
                    ZIP: 20850
          COMPUTER READABLE FORM:
                    MEDIUM TYPE: CD-R
                    COMPUTER: Dell Latitude
                    OPERATING SYSTEM: Windows 98
                    SOFTWARE: ASCII Text
          CURRENT APPLICATION DATA:
                    APPLICATION NUMBER: US/10/206,576
                    FILING DATE: 29-Jul-2002
                    CLASSIFICATION: <Unknown>
          PRIOR APPLICATION DATA:
                    APPLICATION NUMBER: US 09/071,035
                    FILING DATE: 1998-05-04
                    APPLICATION NUMBER: US 60/046,655
                    FILING DATE: 1997-05-16
                    APPLICATION NUMBER: US 60/044,031
                    FILING DATE: 1997-05-06
                    APPLICATION NUMBER: US 60/066,009
                    FILING DATE: 1997-11-14
          ATTORNEY/AGENT INFORMATION:
                    NAME: Hyman, Mark J.
                    REGISTRATION NUMBER: 46,789
                    REFERENCE/DOCKET NUMBER: PB369P1D1
          INFORMATION FOR SEQ ID NO: 258:
                    SEQUENCE CHARACTERISTICS:
                    LENGTH: 1638 amino acids
                    TYPE: amino acid
                    STRANDEDNESS: single
                    TOPOLOGY: linear
                    MOLECULE TYPE: protein
                    SEQUENCE DESCRIPTION: SEQ ID NO: 258:
US-10-206-576-258

Query Match            3.2%;    Score 293;   DB 12;   Length 1638;
Best Local Similarity  21.5%;   Pred. No. 1.6e-12;
Matches  293;   Conservative 140;   Mismatches 499;   Indels 430;   Gaps 72;

Qy   520  SVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAH  579
Db   279  AVGGS---GVNELVDASLLGTTVVT-LPTTVSTPQNL--SNNLDARFVGTVVQTDLLDVN  332

Qy   580  YYWEAYHVLGDGSVMPYRAMPKTNVTLP--ASLPQNQASYSIQASAGSYVAI--SKD    633
Db   333  LLATA----DGVSNIYFAAGTTSEVTAPTITGVTGNSTAGYEVKGTADANATVEIRNAG  387

Qy   634  GVLYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPVIKQIQVGE--PSPYQ-PVS  690
Db   388  GTVIGTGTADGTGAFTV---TVPAGEAGANETLTA----VAKNASGTEXTPTTFQTPAD   439

Qy   691  NLTAT------TQGQKVTLKWEAPSAKKAEGSREVKRI------------GDGLFVTIEPA  733
Db   440  EATVTAPTITGVTGNS-TAGYEVKGTADANATVEIRNAGGTVIGTGTADGTGAFTVTVPA  498

Qy   734  NDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFE  793
Db   499  GEAGANETLTAVAKNASGTESTPTTFQTPADEAT------VTAPTITGVTGNS--TAGYE  550

Db   794  YLVPANADPVVTTQN---IIVTGQGEVVIPGGVYDYCIITNPEPASGKMWIAGDGG----  845
Db   551  VKGTADANATVEIRNAGGAVIGTGTGTAD------GTGAFTVTIP------AGEAGANET  596

Qy   846  ------------------NQPAR---YDDFTFRAGKKYTFMRRAGMGD-GTDM         877
Db   597  LTAVAKNASGTESTPTTFQTPADPNTPVATPIVETVGSTTKGYEV---KGTAEVGTTI    652

Qy   878  EVEDDSPASYTYTVRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVC--   934
Db   653  EVRDAAG------TVLGTATTGTDGKYTVTL-DSGTATANQTLSVVAKNASGTESQPATAT 706

Qy   935  --KDVTVEGSNEFAP-VQNLTGSS-VGQKVTLKWDAPNGTPNPN-----------      974
```

```
Qy  1010  -----------------------------KPGNAPGIAGYNSNGCVYS           1028
Db   634  TPSTPLPNGTVVNATATDPSGNASSPASVTVDAVAPATPVVNPSNGTTLSGTAEPGATVT  693

Qy  1029  ESFGLGG-IGVLTPD--NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGND   1084
Db   694  LTDGNGNPIGQVTADGSGNWSFTPTTPLNG------------TVNATATD           733

Qy  1085  ASNFTNALLEETI---------TAKGVRSPKAIRGRIQGTWR---QKTVDLPA        1125
Db   734  ASGNTSAGSSVTVDSVAPATPVINPSNGTTLSGTAEPGSSVTLTDGNGNPIGQVTAD---  790

Qy  1126  GTKVVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFFS-STHGEAPAEWTTIDADGDG  1184
Db   791  GSGNWSFT--PSTPL-------------ADGTVVNATATDPAGNTSQGS---TTVDGVAPT 834

Qy  1185  QGWLCLSSGQLDWLTAHGGSNVV----------SSFSW-------NGMALNPD        1220
Db   835  TPTVNLSNGSSLSGTAEPGSTVILTDGNGNPIAEVTADGSGNWTYTPSTPIANGTVVN--  892

Qy  1221  NYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGAR  1280
Db   893  --VVAQDAAGNSSPGASVTVDSQAPA--APVVNPSNGT------------TLSGTAEPGAT 937

Qy  1281  FGLSTEANG--------AKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMG 1333
Db   938  VTL-TDGNGNPIGQVTADGSGNW-----SFTPGTPLA-----NGTVVNATASDPTGNTSA  986

Qy  1334  GSPTPTDYTYTVY----RDGTKI---KEGLTETTFEEDG------VATGNHEYCVE     1376
Db   987  PASTTVDSVAPAAPVVNPSNGAEISGTAEPGATVTLLTDGSGNPIGQVTADGSGNWSFTPS 1046

Qy  1377  VKYTAGVSPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFEN  1436
Db  1047  TPIADG-------------TVVNATATDPAGN-TGGQGSTVDAI---APATPVNLSNGSSLS 1093

Qy  1437  GIPASWKT-IDADGDGN--------NWTTTP-----                        1458
Db  1094  GTAEPGSTVILTDGNGNPIAEVTADGSGNWTYTPSTPIANGTVVNVVAQDASGNSSPPAT 1153

Qy  1459  -----PP-----------GGSSFAGHNSAICVSSASHINFEGPQNDNYLVTPELSL     1499
Db  1154  VTVDSSAPPAPVINPSNGVVISGTAEAGATVTLTDAGGNPIGQVTADGSGNWSFTPGTPL  1213

Qy  1500  PGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAI---  1556
Db  1214  ANGTVIV---------ATATDPTGNTGPQAA---------TTVDAVAPPAPVIDP       1250

Qy  1557  -RGTRAQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVVITSGNAPSYTYTI---  1611
Db  1251  SNGTTISGT-------ABAGAK---------------VILTDGNGNPIGETTADGS     1284

Qy  1612  ----YRNNTQIASG-VTETTYRDPDLATGFYTYGVKVVYPNGESAIETATLNIT-----  1660
Db  1285  GNWSFTPGTPLANGTVVNAVAQDPAGNTG---------PQGSTTVDAVAPNTPVVNPSN   1334

Qy  1661  -SLADVTAQCKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNT              1704
Db  1335  GNLLNGTAEP------GSTVTLT---------DGNGNPI--GQTT               1362
```

```
Db   707 TPADVT--------APTVDNITGNSGSGYEIT---------GTADPNTTIEVRDPSGAVIGTG 752
             |||            ||:|:||:||:| |            ||:||
Qy   975 --------------PNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPG---------- 1012
                           |||:|:||:                    |::|
Db   753 TSDANGDFTVTLPTGTTNPGDTLT--------------VIGKDNAGNESQPTEVLVPADATV 800

Qy  1013 NAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASE 1072
             |||:|:|                        :|    |||
Db   801 TAPTVTGVTGNSVA--------------------GYQVTGTAD-PNA---TIEIRDADGN--- 836

Qy  1073 HYAVVASSTGNDASNFT-------NALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAG 1126
             |:|: :::|               ||:||
Db   837 ---VIATGTADGTGSFAVNLPAGTANANETITA-----------------LAKDPAG 873

Qy  1127 TKYVAFRHFQSTDMFYIDLDEV------EIKANGKRA-DFTETFESSTHGEAPAEWTTI 1178
             |:|| ||::::|                 |:|:              |:||
Db   874 --------NTSTPTTFQTPADEVVAPSVDKVTGNTTQGYQVTGTAELGTTIEVRATDGTV 926

Qy  1179 ---DADGDQGWLCLSSGQLDWLTAHGGSNVVSSPSWNGMALNPDNYLLISKDVTGATKV 1234
             |||:|:                  ::|         |||
Db   927 LGTATTGPTGQYTVTLASGK--ATAKQTVNVVAK-NDTGLESQPTTAMTPADVTTPT- 980

Qy  1235 KYYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEET----PNGI---NKGGARFGLSTE 1286
             |||:|                 :|  ||:            |||
Db   981 --------------IGDITGDSTTGYEITGTADPNTTIEVRNPDGTIIGTTTT 1019

Qy  1287 ANGAKPQSVWIERTVDLPAGTKVVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTY--- 1343
             |:|||                          ::|       |:|
Db  1020 DDQG---NFTVDLPAGAA-------NPGDTLTVVGKDGD---GNESQPTEVTVPED 1062

Qy  1344 -TVYRDGTKIKEGLTETTFEBDGVATGNHEYCVEVKYTAG-VSPKKCVDVTVNSTQFNPV 1401
             ||||:|                              :|  :|||
Db  1063 ATVAAPTVTTVTGTTATGYQVTGTABPN--VTIEIHNEAGLVIATGTTDGAGAFTITLPT 1120

Qy  1402 QNLTAEQAPNSMDAILK--------------WNAPASKRAEVLNEDFE------NG---- 1437
             :|                            |:                  |:
Db  1121 GTATANEA---LTAIAKDAAGKESNPTAFKTPADPDAPVATPTVDKITGSTTNGYQVVGA 1177

Qy  1438 -IPASWKTIDADG----------DGNNWTTTPPPGGSSFAGHNSAICVSSASHINFEGP 1485
             |:|                    ||:|:|                  :|
Db  1178 AEVGTTVEVRDADGTVLGMATTGTDG-KYTVTLEFGKAS----ANETITVVAK---NATGK 1230

Qy  1486 QNPDNYLIVTP-ELSLPGGGTLTFWVCAQDANYASEHY----------------AVYA 1525
             ::             ||:|                               |||
Db  1231 ESQPATATTPVDLATPTIDSIT-------GNSSKGYEITGTAEPKTTIDVRDADGTIIA 1282

Qy  1526 SSTGNDASNFANALLEEVLT-AKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHF 1584
             ::|:||                                    :|||
Db  1283 ATTANETGQYTVTLPAGVVTPGETITIISKDGAGNESQ---PATAVIPA--------- 1328

Qy  1585 GCTDFFWINLDDVVITSGNAPSYTYIIYRNNTQIASGVTETTYRDPPLATGFYTYGVKVV 1644
             |::|                                    ::|
Db  1329 --------DVVLA---AP----TITKVEGNKANGYTVTGTADPNVTVQFYNSSEQLL 1370

Qy  1645 YPNGESAIETATLNIISLADVTAQKPYTLTVVGKTIVTCQG 1686
             |::|||   |:|:|            :|:|
Db  1371 ASGNTTTGGTFSVHIA--AGLATEKE-TLT----ALTTDTQG 1405
```

Search completed: May 18, 2004, 11:38:32
Job time : 74.1557 secs

```
              GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        May 18, 2004, 11:33:39 ; Search time 51.2572 Seconds
                                  (without alignments)
                        2739.638 Million cell updates/sec

Title:         US-08-570-311-2
Perfect score: 2641
Sequence:      1 MRKLNLSLFSLAVLLSLLCWG..........QNLTGSAVGQKVTLKWDAPN 497

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5
Searched:      1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:  ,  1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     A_Geneseq_29Jan04.*
            1:  geneseqp1980s:*
            2:  geneseqp1990s:*
            3:  geneseqp2000s:*
            4:  geneseqp2001s:*
            5:  geneseqp2002s:*
            6:  geneseqp2003as:*
            7:  geneseqp2003bs:*
            8:  geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Match	Query Match Length	DB	ID	Description
1	2641	100.0	497	2	AAR96025	Aar96025 P. gingiv
2	2641	100.0	497	2	AAW69483	Aaw69483 Haemagglu
3	2641	100.0	2628	2	AAR96030	Aar96030 P. gingiv
4	2641	100.0	2628	2	AAW69488	Aaw69488 Haemagglu
5	951.5	36.0	1706	2	AAW24786	Aaw24786 PrtR anti
6	944.5	35.8	1704	2	AAR70188	Aar70188 Arg-gingi
7	944.5	35.8	1704	2	AAW34843	Aaw34843 Arg-gingi
8	944.5	35.8	1704	2	AAY67396	Aay67396 Arg-gingi
9	944.5	35.8	1704	4	AAU08938	Aau08938 P. gingiv
10	943.5	35.7	1687	2	AAR96033	Aar96033 P. gingiv
11	943.5	35.7	1687	2	AAW69495	Aaw69495 Haemagglu
12	816	30.9	1358	2	AAR96032	Aar96032 P. gingiv
13	816	30.9	1358	2	AAW69494	Aaw69494 Haemagglu
14	812.5	30.8	1732	2	AAR96029	Aar96029 P. gingiv
15	812.5	30.8	1732	2	AAW24787	Aaw24787 PrtK anti
16	812.5	30.8	1732	2	AAW69487	Aaw69487 Haemagglu
17	806.5	30.5	1087	2	AAR96028	Aar96028 P. gingiv
18	806.5	30.5	1087	2	AAW69486	Aaw69486 Haemagglu
19	791.5	30.0	419	4	AAU03572	Aau03572 P. gingiv
20	662.5	25.1	419	4	AAU03574	Aau03574 P. gingiv
21	546.5	20.7	450	2	AAR96021	Aar96021 P. gingiv
22	546.5	20.7	450	2	AAW69489	Aaw69489 Haemagglu
23	546.5	20.7	456	2	AAR96023	Aar96023 P. gingiv
24	546.5	20.7	456	2	AAR96022	Aar96022 P. gingiv
25	546.5	20.7	456	2	AAW69491	Aaw69491 Haemagglu
26	546.5	20.7	456	2	AAW69490	Aaw69490 Haemagglu
27	522	19.8	970	2	AAR72458	Aar72458 Porphyrom
28	495.5	18.8	991	2	AAR77313	Aar77313 Porphyrom
29	458.5	17.4	439	2	AAR96024	Aar96024 P. gingiv
30	458.5	17.4	439	2	AAW69492	Aaw69492 Haemagglu
31	457.5	17.3	231	4	AAU03575	Aau03575 P. gingiv
32	425.5	16.1	377	2	AAY34359	Aay34359 Porphyrom
33	379	14.4	312	2	AAY34484	Aay34484 Porphyrym
34	337	12.8	921	2	AAY34522	Aay34522 Porphyrym
35	337	12.8	922	2	AAY34521	Aay34521 Porphyrym
36	337	12.8	925	2	AAY34520	Aay34520 Porphyrym
37	337	12.8	938	2	AAY34392	Aay34392 Porphyrym
38	272.5	10.3	148	3	AAB14942	Aab14942 Porphyrom
39	213.5	8.1	293	2	AAY34483	Aay34483 Porphyrym
40	213.5	8.1	299	2	AAY34358	Aay34358 Porphyrym
41	188.5	7.1	196	4	AAU03573	Aau03573 P. gingiv
42	177.5	6.7	135	6	ABP55081	Abp55081 Porphyrom
43	176.5	6.7	134	4	AAB49217	Aab49217 Peptide u
44	175	6.6	49	2	AAW34805	Aaw34805 Arg speci
45	153	5.8	5291	7	ADC01014	Adc01014 Enterohae

ALIGNMENTS

```
RESULT 1
AAR96025
ID   AAR96025 standard; protein; 497 AA.
XX
AC   AAR96025;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;  95WO-US016108.
XX
PR   09-DEC-1994;  94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30649.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 47-51; 153pp; English.
XX
CC   A portion (AAR96025) of P. gingivalis 381 haemagglutinin hagA (see also
CC   AAR96030) was identified as the product of clone ST2 (AAT30649) isolated
CC   from a P. gingivalis 318 genomic library. The haemagglutinin can be obtd.
CC   from transformed host cells and used as a vaccine to protect humans or
CC   animals against periodontal disease. Expression in Salmonella cells
CC   allows prodn. of a live vaccine. The haemagglutinin can also be used to
CC   detect the presence of anti-P. gingivalis antibodies and to raise
CC   monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-2003 to
CC   standardise OS field)
XX
SQ   Sequence 497 AA;
```

```
Query Match             100.0%;  Score 2641;  DB 2;  Length 497;
Best Local Similarity   100.0%;  Pred. No. 2.5e-194;
Matches  497;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60

Qy     61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120

Qy    121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180

Qy    181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240

Qy    241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300

Qy    301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360

Qy    361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420

Qy    421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480

Qy    481 TGSAVGQKVTLKWDAPN 497
            |||||||||||||||||
Db    481 TGSAVGQKVTLKWDAPN 497

RESULT 2
AAW69483
ID   AAW69483 standard; protein; 497 AA.
XX
AC   AAW69483;
XX
DT   22-DEC-1998  (first entry)
XX
DE   Haemagglutinin protein hagA.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;  95US-00570311.
XX
PR   08-SEP-1988;  88US-00241640.
PR   25-JAN-1991;  91US-00647119.
PR   09-DEC-1994;  94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58870.
XX
```

```
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 37-44; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 497 AA;

Query Match             100.0%;  Score 2641;  DB 2;  Length 497;
Best Local Similarity   100.0%;  Pred. No.2.5e-194;
Matches  497;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60

Qy     61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120

Qy    121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180

Qy    181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ 240

Qy    241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP 300

Qy    301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK 360

Qy    361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420

Qy    421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480

Qy    481 TGSAVGQKVTLKWDAPN 497
            |||||||||||||||||
Db    481 TGSAVGQKVTLKWDAPN 497

RESULT 3
AAR96030
ID   AAR96030 standard; protein; 2628 AA.
XX
AC   AAR96030;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
FH   Key            Location/Qualifiers
FT   Peptide        5..21
FT                  /label= Sig_peptide
```

```
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30654.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 6; Page 93-101; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagA (AAR96030) was identified as the
CC   product of the hagA gene (AAT30654) isolated as an EcoRV fragment of
CC   genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
CC   24), can be obtd. from transformed host cells and used as a vaccine to
CC   protect humans or animals against periodontal disease. Expression in
CC   Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
CC   also be used to detect the presence of anti-P. gingivalis antibodies and
CC   to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-
CC   2003 to standardise OS field)
XX
SQ   Sequence 2628 AA;

     Query Match          100.0%;      Score 2641;  DB 2;    Length 2628;
     Best Local Similarity 100.0%;      Pred. No.2.9e-193;
     Matches 497;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy      1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60

Qy     61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120

Qy    121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180

Qy    181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240

Qy    241  TLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP   300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  TLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP   300

Qy    301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK  360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK  360

Qy    361  SEGVKLPAPYQERTIDLSAYAGQQYYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  SEGVKLPAPYQERTIDLSAYAGQQYYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420

Qy    421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480

Qy    481  TGSAVGQKVTLKWDAPN  497
           |||||||||||||||||
Db    481  TGSAVGQKVTLKWDAPN  497
```

```
RESULT 4
AAW69488
ID   AAW69488 standard; protein; 2628 AA.
XX
AC   AAW69488;
XX
DT   22-DEC-1998   (first entry)
XX
DE   Haemagglutinin protein hagA.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;     95US-00570311.
XX
PR   08-SEP-1988;     88US-00241640.
PR   25-JAN-1991;     91US-00647119.
PR   09-DEC-1994;     94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58875.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 91-110; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 2628 AA;

     Query Match          100.0%;      Score 2641;   DB 2;    Length 2628;
     Best Local Similarity 100.0%;      Pred. No. 2.9e-193;
     Matches 497;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy      1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60

Qy     61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120

Qy    121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180

Qy    181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240

Qy    241  TLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP   300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  TLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP   300

Qy    301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK  360
```

```
PT   New antigenic protein complex from Porphyromonas gingivalis - comprising
PT   Arg- and Lys- specific thiol endo-peptidase(s), used in the detection,
PT   prevention and treatment of periodontal disease.
XX
PS   Example 1; Fig 8b; 68pp; English.
XX
CC   A PrtR-PrtK cell surface protein of Porphyromonas ginivalis (PG)
CC   comprises a 300 kDa complex composed a 45 kDa arginine-specific thiol
CC   protease and 44, 15, 17 and 27 kDa adhesins encoded by the prtR gene
CC   (AAT78850), and a 148 kDa lysine-specific thiol protease and 39, 15 and
CC   44 kDa adhesins (see AAW24787) encoded by the prtK gene (AAT78851). A
CC   claimed antigenic complex comprises at least one multimeric protein
CC   complex of PrtR and PrtK each containing at least one adhesin domain, the
CC   complex having a mol. wt. of over 200 kDa, and preferably comprises all 9
CC   proteins of the Prt-PrtK complex (see also AAW24780-85). It can be used
CC   in a claimed composition to elicit an immune response directed against
CC   PG, and in a claimed method of reducing the prospect of PG infection
CC   and/or severity of disease. Antibodies directed against the complex are
CC   claimed for use in treating PG infection. Unlike whole PG cells or other
CC   previously prepared antigens based on fimbriae or the capsule, the PrtR-
CC   PrtK complex or component parts are safe and effective antigens. (Updated
CC   on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1706 AA;

Query Match           36.0%;    Score 951.5;   DB 2;    Length 1706;
Best Local Similarity 31.9%;    Pred. No. 1.3e-63;
Matches 237;   Conservative 74;   Mismatches 156;   Indels 275;   Gaps   20;

Qy    22 TAAAQG-------------------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
         |||||                                                     :|:|||:|||
Db   671 TATTQQGKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy    69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
         |||||||::||||:|| ||:|||| :: |: :|| :::::|
Db   731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
         ||:.|||||:|||||||||:|::::||  |  :  |:|||:  :|| :||||:::|:|||:
Db   790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy   186 PGDAASVV-- 193
         ||   :
Db   850 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909

Qy   194 -------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------------DKRTYVLNE 235
         ||:||||
Db   910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSE 969

Qy   236 SFDTQTLPNGWTMIDADGDGHNWLSTNVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
         ||: :  :|
Db   970 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1025

Qy   296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETL--GSD 351
         |||:|| ||:|:|  :|||:|:: :|:|||||:||:|:|:||:|||:|:||||
Db  1026 YLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1085

Qy   352 KPAPM---------------- 356
         ||:|
Db  1086 SPEAMRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1145

Qy   357 -----------------NLVKS------------ 361
         :|:|
Db  1146 ESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALNPDNYL 1205

Qy   362 362
Db  1206 ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGL 1265

Qy   362 ---EGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDV--AVSGEGSSND 415
         :  :: |:::|  :|:|:|:|:|:  : :|||
Db  1266 STEADGAKPQSVWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD 1322
```

```
Db   301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDKPAPMNLVK 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
Qy   421 YRDNVVIAQNLAATFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 YRDNVVIAQNLAATFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
Qy   481 TGSAVGQKVTLKWDAPN 497
         |||||||||||||||||
Db   481 TGSAVGQKVTLKWDAPN 497

RESULT 5
AAW24786
ID   AAW24786 standard; protein; 1706 AA.
XX
AC   AAW24786;
XX
DT   17-OCT-2003  (revised)
DT   25-NOV-1997  (first entry)
XX
DE   PrtR antigenic protein complex.
XX
KW   Periodontal disease; cell surface protein; thiol protease; endopeptidase;
KW   PrtR; PrtR45; PrtR44; PrtR15; PrtR17; PrtR27; haemagglutinin; adhesin;
KW   therapy; diagnosis; vaccine; antigen.
XX
OS   Porphyromonas gingivalis; strain W50.
XX
FH   Key         Location/Qualifiers
FT   Peptide     1..227
FT               /label= Pro-pro_peptide
FT   Cleavage-site  227..228
FT   Protein     228..719
FT               /label= PrtR45
FT               /note= "45 kDa Arg-specific thiol protease"
FT   Cleavage-site  719..720
FT   Protein     720..1138
FT               /label= PrtR44
FT               /note= "44 kDa adhesin"
FT   Cleavage-site  1138..1139
FT   Protein     1139..1273
FT               /label= PrtR15
FT               /note= "15 kDa adhesin"
FT   Cleavage-site  1273..1274
FT   Protein     1274..1431
FT               /label= PrtR17
FT               /note= "17 kDa adhesin"
FT   Cleavage-site  1431..1432
FT   Protein     1432..1706
FT               /label= PrtR27
FT               /note= "27 kDa adhesin"
XX
PN   WO9716542-A1.
XX
PD   09-MAY-1997.
XX
PF   30-OCT-1996; 96WO-AU000673.
XX
PR   30-OCT-1995; 95AU-00006275.
XX
PA   (UYME ) UNIV MELBOURNE.
PA   (VICT-) VICTORIAN DAIRY IND AUTHORITY.
XX
PI   Reynolds EC,  Bhogal PS,  Slakeski N;
XX
DR   WPI; 1997-272112/24.
DR   N-PSDB; AAT78850.
XX
```

```
Qy   22 TAAAQG-------GPKTAPSVTHQAVQK--GIR--TSKVKDLRDPIPAGMARIIIEAH   68
                      ||            :|       |||    :  :|:  :||
Db  671 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELLVLLSVSDAPELLRSGQAEIVLEAH  730

Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP  126
         ||:|  ||: |:|||||:||| | |:    : |   || || | :|||:|| |||  |:|
Db  731 DVWNDSGYQILLDADHDQYGVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP    789

Qy  127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG  185
         || ::|||| ||||||||:.|: ||.| | ||.|  |: ::||| |||||||||  :|:|
Db  790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG  849

Qy  186 PGDAASVV------                                               193
         |:||::
Db  850 SGDGTELTISGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS   909

Qy  194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------DKRTYVLNESF     237
               :| |.||||||||||  |.| |: | |:| || |||        :|:| | ||:|
Db  910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESF   969

Qy  238 DTQTLPNGWTMIDADGGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL   297
          : | ||.| :|||||.|..| | |||:|::||.  |  .|.  ||.||::|:|:|.|||
Db  970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy  298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETIGS------  350
          ||  | |:|||| :||  |  |   |||| : |:||:.|:||| |||||| |
Db 1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAYYASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy  351 ---------DKPAPMNLV------------KSEGVKL-                       366
                   :|.|||:|             |||.|:|
Db 1086 EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145

Qy  367 ---PAPYQERTID------------LSAYAGQQV------                     385
            .||:|||: ||            |:||.:
Db 1146 STHGEAPAEWTTIDADGDGGWLCLSSGGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy  386 ---------------                                              385
Db 1206 KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGLST  1265

Qy  386 --------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV          420
                       ||||||| |:||| || |:| ||  ||| |.|||:|||||
Db 1266 EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNVILLDDIQFTMGGSPTPTDYYTYTV 1325

Qy  421 YRDNVVIAQNLAATFNQENVAPGQYNYCVEVKYIAGVSPKCVCKDVTVEGSNEFAHVQNL  480
         |||.| |:: |.||| ||.|| |: :||||||||.|||||.||:||||||||||||||:||
Db 1326 YRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL  1384

Qy  481 TGSAVGQKVTLKWDAPN                                            497
         :|:|:|.|||||.|||:
Db 1385 KAQPDGGDVVLKWEAPS                                           1401
```

```
Qy  416 YTYTVYRDNVVIAQNLAATFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA  475
         ||||||||:. |:: |.||| ||.|| || :||||||||||||||:.||:||| ||.|
Db 1323 YTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN 1381

Qy  476 HVQNLTGSAVGQKVTLKWDAPN                                       497
          |:|| |:|:|.|||||.|||:
Db 1382 PVKNLKAQPDGGDVVLKWEAPS                                      1403
```

RESULT 6
AAR70188
ID   AAR70188 standard; protein; 1704 AA.
XX
AC   AAR70188;
XX
DT   25-MAR-2003 (revised)
DT   21-SEP-1995 (first entry)
XX
DE   Arg-gingipain-2 prepolyprotein.
XX
KW   Arg-gingipain-2; gingivalis; periodontal disease; vaccine;
KW   arginine-specific protease.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key   Location/Qualifiers
FT   Protein   228..719
FT   /label= Protease
FT   /note= "corresponds to Arg-gingipain-1"
FT   Region   720..1091
FT   /label= Hemagglutinin
FT   Region   1092..1429
FT   /label= Hemagglutinin
FT   Region   1430..1704
FT   /label= Hemagglutinin
XX
PN   WO9507286-A1.
XX
PD   16-MAR-1995.
XX
PF   09-SEP-1994; 94WO-US010283.
XX
PR   10-SEP-1993; 93US-00119361.
PR   21-OCT-1993; 93US-00141324.
PR   24-JUN-1994; 94US-00265441.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J, Potempa J, Barr PJ, Pavloff N;
XX
DR   WPI; 1995-123373/16.
DR   N-PSDB; AAQ83489.
XX
PT   DNA encoding Arg-gingipain proteins - used to develop prods. for
PT   detection, treatment and prevention of periodontal disease.
XX
PS   Disclosure; Page 70-77; 89pp; English.
XX
CC   A low mol.wt. arginine-specific gingipain (AG-1) and high mol.wt. AG (AG-
CC   2) were isolated from P. gingivalis strains H66 (ATCC 33277) and W50
CC   (ATCC 53973). The sequences of the proteins were used to design PCR
CC   primers and probes to isolate AG DNA. Lambda DASH and lambda ZAP
CC   libraries were screened with a probe based on amino acids 11-22 of the AG
CC   protein to obtain DNA encoding AG-1 (AAQ83484) and AG-2 (AAQ83489). AG-2
CC   is a prepolyprotein incorporating AG-1. (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 1704 AA;

Query Match   35.8%; Score 944.5; DB 2; Length 1704;
Best Local Similarity   31.9%; Pred. No. 4.4e-63;
Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

RESULT 7
AAW34843
ID   AAW34843 standard; protein; 1704 AA.
XX
AC   AAW34843;
XX
DT   03-JUN-1998 (first entry)
XX
DE   Arg-gingipain high molecular weight prepolyprotein sequence.
XX
KW   Arg-specific gingipain protease; gingivalis; periodontal disease;
KW   vaccine; infection.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key   Location/Qualifiers
FT   Protein   1..227
FT   /note= "precursor protein"
XX
PN   WO9734629-A1.

```
Db  1146  STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS  1205
                                                  ||:|: |
Qy   386  ----------------------------------------------------------  385

Db  1206  KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGLST  1265
Qy   386  -------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV  420
                              :||||| : :: ||| : |||:: | |:   ||||

Db  1266  EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYTYTV  1325
                           |:|||||||||||| :| :::  : :||    | | ||||||
Qy   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
          :|| |:||||||||||||||||||||||||||||||||||| ||||:| ||||    ||

Db  1326  YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL  1384
Qy   481  TGSAVGQKVTLKWDAPN  497
          | |: |:|||||| :||
Db  1385  KAQPDGGDVVLKWEAPS  1401
```

```
RESULT 8
AAY67396
ID   AAY67396 standard; protein; 1704 AA.
XX
AC   AAY67396;
XX
DT   25-APR-2000  (first entry)
XX
DE   Arg-gingipain-2 amino acid sequence.
XX
KW   Arginine specific proteinase; Arg-gingipain; gingipain-2; haemagglutinin;
KW   immunogenic component; vaccine; inflammatory response; tissue damage;
KW   periodontal disease.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key            Location/Qualifiers
FT   Region         229..719
FT                  /note="Amino acids 229-719 are specifically claimed"
FT   Region         720..1185
FT                  /note="Amino acids 720-1185 are specifically claimed"
XX
PN   US6017532-A.
XX
PD   25-JAN-2000.
XX
PF   08-NOV-1994;  94US-00336308.
XX
PR   10-SEP-1993;  93US-00119361.
PR   24-JUN-1994;  94US-00265441.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Potempa JS, Travis J;
XX
DR   WPI; 2000-136659/12.
DR   N-PSDB; AAZ60181.
XX
PT   New Porphyromonas gingivalis arginine-specific protease preparation
PT   useful for preparing vaccines against periodontal disease and for
PT   screening for Arg-gingipain inhibitors.
XX
PS   Claim 1; Col 29-42; 55pp; English.
XX
CC   This sequence represents a Porphyromonas gingivalis arginine-specific
CC   proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
CC   Gingipain-2 consists or a 50kD protease component non-covalently
CC   associated with a 44kD haemagglutinin component. The proteinase is
CC   stimulated by glycine containing peptides and glycine analogues. It is
CC   inhibited by cysteine protease group specific inhibitors. The protease
CC   preparation can be used in immunogenic compositions and vaccines against
CC   inflammatory response and tissue damage caused by P. gingivalis in
CC   periodontal disease. It can also be used to screen for agents that
```

```
XX
PD   25-SEP-1997.
XX
PF   21-MAR-1997;  97WO-US004635.
XX
PR   22-MAR-1996;  96US-0013945P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
PA   (MORE-) MOREHOUSE SCHOOL MEDICINE.
XX
PI   Potempa J, Travis J, Genco C;
XX
DR   WPI; 1997-479993/44.
DR   N-PSDB; AAT93872.
XX
PT   Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
PT   useful for protecting animals and humans from gingivalis and periodontal
PT   diseases.
XX
PS   Disclosure; Page 68-73; 95pp; English.
XX
CC   The present sequence represents an arginine-specific protease of
CC   Porphyromonas gingivalis. The following peptides, derived from Arg- and
CC   Lys-specific high molecular weight proteases, offer protection against
CC   infection: YTYTVVRDGK IKEGLTATTE DDGVATGNHE YCVEKYTAGS VSPKVC (I);
CC   YTPVEEKQNG RMIVIVAKKY  (II); QLPFIPDVAC VNGDFLFSMP CFAEALMRAQ (III);
CC   GEPNPYQPVS NLTATTQGGK VTLKWDAPSTK (IV); GNHEYCVEVK YTAGVSPKVC KDVTV (V);
CC   RMFMNYEPGR YTPVEEKQNG  (VI); TFAGFEDYTK RMFMNYEPGR (VII); DYTYTVYRDG
CC   TKIKEGLTAT TFEEDGVATG NMEYCVCVKY TAGVSPKVC (VIII); YTYTVVRDGT KIKEGLTATTF
CC   EEDG (IX); RDGTKIKEGL TATTFEEDGV ATGN (X); KIKEGLTATT FEEDGVATGN HEY (XI)
CC   ; KWDAPNGTPN PNPN PNPNPN PGTTTLSE (XII); and YTPVEBKENG RMIVIVAKKY
CC   (XIII). They are used in vaccines to protect animals, including humans,
CC   from gingivitis and/or periodontal diseases
XX
SQ   Sequence 1704 AA;
```

```
Query Match      35.8%;  Score 944.5;  DB 2;  Length 1704;
Best Local Similarity  31.9%;  Pred. No. 4.4e-63;
Matches  235;  Conservative 75;  Mismatches 160;  Indels 267;  Gaps  20;

Qy    22  TAAAQG-------------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH  68
          |||  | |:|||   : |:|||| :|  :::    |: ||:||||
Db   671  TATTQGGKVTLKWDAPSTKTNATTNTARSVDGIRELIVLLSVSDAPELLRSGQAEIVLEAH  730

Qy    69  DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP  126
          |||:|||::||:::|||| : :|||    ::::| |  ||||||| |||||  :|
Db   731  DVWNDGSGYQILLDADHDQYGGVIPSDTHTLW PNCSVPANLFAPFEYTVPENADFSCSP  789

Qy   127  TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG  185
          ||  :||||:|:|:|  :: : ||| |: |:||||| :| |||: ::|:||:::
Db   790  TNMIMDGTASVNIPAGTTDFAIIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG  849

Qy   186  PGDAASVV--  193
          ||
Db   850  SGDGTELITISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS  909

Qy   194  ----VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS--------DKRTYVLNESF  237
          |   ||:|||:|||||||| |  |: :|  ||:|       |: |:|||
Db   910  PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESF  969

Qy   238  DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL  297
          |||||   :|:||||||||: |::|:|  ::  |||| || ||:||||||||
Db   970  E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL  1025

Qy   298  VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVPLSTTGNEAANFTIKLLEEFLGS------  350
          :|:|: |:|  |  |:::   | ::||  |:  |:  ||:||:|::|||||
Db  1026  ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP  1085

Qy   351  -----------DKPAPMNLV-------------KSEGVKL--------  366
           |:|  |||         |:|
Db  1086  EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES  1145

Qy   367  ----PAPYQERTID-----------------LSAYAGQQV---------  385
```

```
CC   modulate Arg-gingipain proteinase activity inhibitors
XX
SQ   Sequence 1704 AA;

Query Match            35.8%;  Score 944.5;  DB 3;  Length 1704;
Best Local Similarity  31.9%;  Pred. No. 4.4e-63;
Matches 235;  Conservative 75;  Mismatches 160;  Indels 267;  Gaps 20;

Qy    22 TAAAQG------GPKTAPSVTHQAVQK--GIR--TSKVKDLRDPIPAGMARIILEAH 68
                : ||:.||:||:||          :    :.:|:|||||  ||||
Db   671 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIREIVLLSVSDAPELLRSGQAEIVLEAH 730

Qy    69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
            ||:|:||:|:||||:|:|          ::||||||||||||||||||
Db   731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAFFEYTVPENADPSCSP 789

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
              |.:: |||:|:||||||||        : :::  ::: ::  || ::|: ||:
Db   790 TNMIMDGTASVNIPAGTYDYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy   186 PGDAASVV-- 193
            ||: |
Db   850 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909

Qy   194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------------DKRTYVLNESF 237
               ||||:| ||:||:|||||       |:|||| |  :  .|:|:||:|
Db   910 PKVCKDVTVEVGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESF 969

Qy   238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
            :  : :|||   ||:|:|           ||:||
Db   970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy   298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAAANFTIKLLEETLGS----- 350
          :||| || ::|  | ||:: |       |||:|::::|     ::|::  ::|
Db  1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETIITAKGVRSP 1085

Qy   351 ----------DKPAPMNLV-- 366
              ||:|:|||:
Db  1086 EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145

Qy   367 ---PAPYQERTID------------LSAYAGQQV--------------- 385
              ||:|:|:||            |:.|  ||:|
Db  1146 STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy   386 ---------------- 386

Db  1206 KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTGNAGDFTVVFEETPNGINKGGARFGLST 1265

Qy   386 ------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
                          :|:||||:| : :| |:||:  | :|| :
Db  1266 EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTV 1325

Qy   421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
           |:|| :| |:: |: :::|||  :|||:||||||||||| : :::|||||:||||  |::
Db  1326 YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1384

Qy   481 TGSAVGQKVTLKWDAPN 497
           ||| ||||| |||: |:
Db  1385 KAQPDGGDVVLKWEAPS 1401
```

```
RESULT 9
AAU08938
ID   AAU08938 standard; protein; 1704 AA.
XX
AC   AAU08938;
XX
DT   18-DEC-2001  (first entry)
XX
DE   P. gingivalis high molecular weight Arg-gingipain-2.
XX
KW   Periodontitis; antiinflammatory; Arg-gingipain-2; AG-2; immunogen.
XX
```

```
OS   Porphyromonas gingivalis.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..227
FT                   /label= Prepro_peptide
FT   Region          599..619
FT                   /note="Region of homology with cysteine proteases"
FT   Region          670..674
FT   Cleavage-site   719
FT                   /label= Proteolytic_component
FT   Protein         720..1091
FT                   /label= HGP 44kDa
FT                   /note="Haemagglutinin protein component"
FT   Cleavage-site   1091
FT   Protein         1092..1429
FT                   /label= HGP 17kDa
FT                   /note="Haemagglutinin protein component"
FT   Cleavage-site   1429
FT   Protein         1430..1704
FT                   /label= HGP 27kDa
FT                   /note="Haemagglutinin protein component"
XX
PN   US6274718-B1.
XX
PD   14-AUG-2001.
XX
PF   25-JAN-2000;  2000US-00490931.
XX
PR   10-SEP-1993;  93US-00119361.
PR   24-JUN-1994;  94US-00265441.
PR   09-SEP-1994;  94WO-US010283.
PR   08-NOV-1994;  94US-00336308.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Barr PJ,  Pavloff N;
XX
DR   WPI; 2001-588904/66.
DR   N-PSDB; AAS15242.
XX
PT   New recombinant DNA molecule which encodes high molecular weight (mature)
PT   Arg-gingipain protein, useful for immunization against inflammation and
PT   tissue damage, comprises enzymatically active protease component and
PT   hemagglutinin component.
XX
PS   Claim 1; Col 29-41; 56pp; English.
XX
CC   The invention relates to a recombinant DNA molecule encoding high
CC   molecular weight (mature) Arg-gingipain (AG) protein, which has an
CC   enzymatically active protease component (AG-2) and a haemagglutinin
CC   component, from P. gingivalis. The nucleic acid is useful for producing
CC   mature Arg-gingipain protein. Immunogenic compositions comprising Arg-
CC   gingipain are useful for immunising animals including humans against
CC   inflammatory response and tissue damage caused by an archaebacterium
CC   Porphyromonas gingivalis, which causes progressive periodontitis. Arg-
CC   gingipain is also useful for identifying agents that modulate Arg-
CC   gingipain proteinase activity, whether by acting on the proteinase itself
CC   or preventing the interaction of the proteinase with the protein in the
CC   gingival area, such as complement factors C3 or C5. The present sequence
CC   is Arg-gingipain-2
XX
SQ   Sequence 1704 AA;

Query Match            35.8%;  Score 944.5;  DB 4;  Length 1704;
Best Local Similarity  31.9%;  Pred. No. 4.4e-63;
Matches 235;  Conservative 75;  Mismatches 160;  Indels 267;  Gaps 20;

Qy    22 TAAAQG------GPKTAPSVTHQAVQK--GIR--TSKVKDLRDPIPAGMARIILEAH 68
                : ||:.||:||:||          :    :.:|:|||||  ||||
Db   671 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIREIVLLSVSDAPELLRSGQAEIVLEAH 730

Qy    69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
            ||:|:||:|:||||:|:|          ::||||||||||||||||||
```

PI   Proguske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30656.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 138-143; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagE (AAR96033) was identified as the
CC   product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
CC   The haemagglutinin can be obtd. from transformed host cells and used as a
CC   vaccine to protect humans or animals against periodontal disease.
CC   Expression in Salmonella cells allows prodn. of a live vaccine. The
CC   haemagglutinin can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1687 AA;

Query Match   35.7%;   Score 943.5;   DB 2;   Length 1687;
Best Local Similarity   31.8%;   Pred. No. 5.2e-63;
Matches 235;   Conservative   75;   Mismatches 160;   Indels 269;   Gaps   20;

Qy   22 TAAAQG---------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
                                :   .| :   : |: .:|:||
Db   652 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 711

Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
               :||:|| : :||||:|||:| || :   .||||:||||| ||:||||||:|| ||||
Db   712 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 770

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
                |.|:|||||:.||||||||:|||||||| ||:||||||.:|||:||||||||.||.:||
Db   771 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGGQGPTKEDDYVFEAGKKYHFLMKKMG 830

Qy   186 PGDAASVV--------   193
               :
Db   831 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 890

Qy   194 --------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS-------------------- 235
                       |:|||||||||||||||:||||||:||.|||||
Db   891 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPNPNPGTTTLSE 950

Qy   236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
                    : :||.||||||:|:|||: :|:| ||:|:|||||. |: ..|||.:|:|:||
Db   951 SFB-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1006

Qy   296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
               ||:||: |||| :: ||:| : ||::|.   :|::||| |:|||: .
Db   1007 YLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETTAKGVR 1066

Qy   351 -------DKPAPMNLV----------------------------KSEGVKL------- 366
                      |||||||| :||| ||:
Db   1067 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLLDEVEIKANGKRADFTETF 1126

Qy   367 -----PAPYQERTID---   385
                    |||||||||
Db   1127 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL 1186

Qy   386 LSAYAGQQV-----------   385
               |:|:|||||
Db   1187 ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGL 1246

Qy   386 ------------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTV 418
                                       ||:||||:|||||||:||||| |:|||||.:||||
Db   1247 STEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTY 1306

Qy   419 TVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ 478
               ||||| :|:|||||||.||| |||||||||||||||||||||:| |||:|||||||:|
Db   1307 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVK 1365

Qy   479 NLTGSAVGQKVTLKWDAPN 497
          || |||:||:
Db   1366 NLKAQPDGGDVVLKWEAPS 1384

RESULT 11
AAW69495
ID   AAW69495 standard; protein; 1687 AA.
XX
AC   AAW69495;
XX
DT   22-DEC-1998   (first entry)
XX
DE   Haemagglutinin protein hagE.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagE.
XX
OS   Porphyromonas gingivalis.
XX
PN   U55824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58881.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 167-182; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagE haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1687 AA;

Query Match   35.7%;   Score 943.5;   DB 2;   Length 1687;
Best Local Similarity   31.8%;   Pred. No. 5.2e-63;
Matches 235; Conservative   75; Mismatches 160; Indels 269; Gaps   20;

Qy   22 TAAAQG-------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
          || : : :| | : ::: |:| |||||
Db   652 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 711

Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKYPVNADASFSP 126
          ||:|||:||||||:|||||||| | | :: |:||||| ||||||||||
Db   712 DVWNDGSGYQILLDADHDQYGQVIPSDTHTIW-PNCSVPANLFAPFEYTVPENADPSCSP 770

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGBGVSKGNDYVVEAGKTYHFTVQRQG 185
          :| ||||| |||||||||| ||||:| :|| ||| :|:: :||||| |||:
Db   771 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDVVFEAGKKYHFLMKKWG 830

Qy   186 PGDAASVV------- 193
          |||:|:||
Db   831 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 890

Qy   194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS-------------DKRTYVLNE 235
          || ||||||| : | || : | | : :|

```
CC   derived from P. gingivalis 318 genomic DNA. A first open reading frame
CC   coded for hagD protease (see also AAR96031). The protease and
CC   haemagglutinin can be obtd. from transformed host cells and used in
CC   vaccines to protect humans or animals against periodontal disease.
CC   Expression in Salmonella cells allows prodn. of live vaccines. The
CC   haemagglutinin and protease can also be used to detect the presence of
CC   anti-P. gingivalis antibodies and to raise monoclonal antibodies for
CC   diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1358 AA;

Query Match             30.9%;  Score 816;  DB 2;  Length 1358;
Best Local Similarity   29.6%;  Pred. No. 2.4e-53;
Matches 219;  Conservative  75;  Mismatches 171;  Indels 274;  Gaps   21;

Qy      22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARIILEA  67
               ||||                  ||  :  ||            |:||||||||||
Db     328 TATTQGQKVTLKWDAPS----AKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAA 382

Qy      68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
            |||||||  ||  ||||||  ||||||:  ||||| ||| |  |  |||  |||||  ||
Db     383 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 441

Qy     127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
            ::  ||| |||   | |  |  |||  || :|| |: :   || ||| |||  || |||
Db     442 QNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGDNQPARYDDFTFEAGKKYTFTM 501

Qy     182 QRQGPGDAASVV--------------------------------------------  193
            :||| ||  |:
Db     502 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 561

Qy     194 -------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTYVLNE 235
                         |||||  |||| |  | :|:  || ||  ||           |||||||
Db     562 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSE 621

Qy     236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
            |:    |  || |||||| |  | |:||  |: | ||   |  |  |||  | |  ||
Db     622 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 677

Qy     296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS----- 350
            | ||  :|:  ||:|  |||:  :||| || |||||| || ||||||  || ||:  :
Db     678 YLITPAIDLANGGKLITFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 737

Qy     351 ---------------DKPAPMNLV------------------------KSEGVKL------  366
                           :|  |  |:                       |  |:|
Db     738 SPEAIRGRIQGTWRQKTVDLPAGTKVVAFRHFQSTMFYIDLDEVEIKANGKRADFTETF  797

Qy     367 ----PAPYQERTID-----------------------LSAYAGQQV--------------  385
                || |  | |                       | | |||:
Db     798 ESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNFDNYL  857

Qy     386 ------------------------------------------------------------  385
Db     858 ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGL  917

Qy     386 -------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTY-------------  418
                         || |||| |||| ||| |||   ||| |||::| :|
Db     918 STEANGAKPQSVWIERTVDLPAGTKVVAFRHYNCSDLDYILLDDIQFTMGGSPTPTDYTY  977

Qy     419 TVYRDNVVIAQNLAATTFNQENVAPGQVNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ  478
                ||:| ||  | |  ||:    ||  | ||||||||||||||||||||     || |
Db     978 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNPVK 1036

Qy     479 NLTGSAVGQKVTLKWDAPN 497
            |  | ||  ||| || ||:
Db    1037 NLKAQPDGDVVLKWEAPS 1055

RESULT 13
AAW69494
ID  AAW69494 standard; protein; 1358 AA.
XX
AC  AAW69494;
```

```
XX   22-DEC-1998  (first entry)
DT
XX   Haemagglutinin protein hagD.
DE
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagD.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;    95US-00570311.
XX
PR   08-SEP-1988;    88US-00241640.
PR   25-JAN-1991;    91US-00647119.
PR   09-DEC-1994;    94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58880.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 145-158; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagD haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1358 AA;

Query Match             30.9%;  Score 816;  DB 2;  Length 1358;
Best Local Similarity   29.6%;  Pred. No. 2.4e-53;
Matches 219;  Conservative  75;  Mismatches 171;  Indels 274;  Gaps   21;

Qy      22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARIILEA  67
               ||||                  ||  :  ||            |:||||||||||
Db     328 TATTQGQKVTLKWDAPS----AKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAA 382

Qy      68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
            |||||||  ||  ||||||  ||||||:  ||||| ||| |  |  |||  |||||  ||
Db     383 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 441

Qy     127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
            ::  ||| |||   | |  |  |||  || :|| |: :   || ||| |||  || |||
Db     442 QNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGDNQPARYDDFTFEAGKKYTFTM 501

Qy     182 QRQGPGDAASVV--------------------------------------------  193
            :||| ||  |:
Db     502 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 561

Qy     194 -------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS--------DKRTYVLNE 235
                         |||||  |||| |  | :|:  || ||  ||          |||||||
Db     562 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSE 621

Qy     236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
            |:    |  || |||||| |  | |:||  |: | ||   |  |  |||  | |  ||
Db     622 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 677

Qy     296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS----- 350
            | ||  :|:  ||:|  |||:  :||| || |||||| || ||||||  || ||:  :
Db     678 YLITPAIDLANGGKLITFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 737
```

```
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30653.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 76-81; 153pp; English.
XX
CC   P. gingivalis W12 cysteine protease, porphypain (AAR96029), was
CC   identified as the product of the prtP gene (AAT30653) isolated from P.
CC   gingivalis W12 genomic DNA. The porphypain shows homology to the
CC   haemagglutinins (see also AAR96026-28 and AAR96030-33) of P. gingivalis
CC   318. It can be obtd. from transformed host cells and used as a vaccine to
CC   protect humans or animals against periodontal disease. Expression in
CC   Salmonella cells allows prodn. of a live vaccine. The porphypain and
CC   haemagglutinins can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1732 AA;

Query Match             30.8%;  Score 812.5;  DB 2;  Length 1732;
Best Local Similarity   30.0%;  Pred. No. 6.4e-53;
Matches  222;  Conservative  70;  Mismatches 172;  Indels 277;  Gaps   22;

Qy   22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD-----------PIPAGMARIILEA 67
Db  693 TATTQGGKVTLKWEAPS----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy   68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
Db  748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVTT 806

Qy  127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
Db  807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGNQPARYDFTFEAGKKYTFTM 866

Qy  182 QRQGPGDAASVV------------------------ 193
Db  867 RRAGMGDGTDMEVEDDSPASYTYTVRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy  194 --------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP-------------ASDKRTYVLN 234
Db  927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTNPNPNPNPGTTLS 986

Qy  235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMPSKSWTASGGAKIDLSPD 294
Db  987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042

Qy  295 NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAAANFTIKLLEETLGG--- 350
Db 1043 NYLITPALDLPNGGKLJTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102

Qy  351 --------------DKPAPMNLV-------------------KSBGVKL 366
Db 1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy  367 -------PAPYQERTID--------------LSAYAGQQV 385
Db 1163 FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy  386 --------------- 385
Db 1223 LISKDVTGATKVKYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEETPNGINKGGARFG 1282

Qy  386 ------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
Db 1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT 1342

Qy  418 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
Db 1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTV-NSTQFNPV 1401
```

```
Qy  351 --------------DKPAPMNLV-------------------KSBGVKL------ 366
Db  738 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 797

Qy  367 -------PAPYQERTID--------------LSAYAGQQV---------- 385
Db  798 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL 857

Qy  386 --------------- 385
Db  858 ISKDVTGATKVKYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEETPNGINKGGARFGL 917

Qy  386 ------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 418
Db  918 STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYILLDDIQFTMGGSPTPTDYTY 977

Qy  419 TVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ 478
Db  978 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNPVK 1036

Qy  479 NLTGSAVGQKVTLKWDAPN 497
Db 1037 NLKAQPDGGDVVLKWEAPS 1055

RESULT 14
AAR96029
ID   AAR96029 standard; protein; 1732 AA.
XX
AC   AAR96029;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis porphypain.
XX
KW   Porphypain; haemagglutinin; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain W12.
XX
FH   Key            Location/Qualifiers
FT   Region         688..708
FT                  /note= "Pro-Asn repeat region type 1"
FT   Region         887..952
FT                  /note= "Pro-Asn repeat region type 2"
FT   Region         946..967
FT                  /note= "Pro-Asn repeat region type 1"
FT   Region         985..1006
FT                  /note= "Pro-Asn repeat region type 3"
FT   Region         1041..1100
FT                  /note= "Pro-Asn repeat region type 4"
FT   Region         1341..1405
FT                  /note= "Pro-Asn repeat region type 2"
FT   Region         1430..1451
FT                  /note= "Pro-Asn repeat region type 3"
FT   Region         1488..1547
FT                  /note= "Pro-Asn repeat region type 4"
FT   Region         1607..1650
FT                  /note= "Pro-Asn repeat region type 2"
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;  95WO-US016108.
XX
PR   09-DEC-1994;  94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
```

```
Qy    478  QNLTGSAV--GQKVTLKWDAP 496
                ||||:|| ||:||
Db   1402  QNLTAEQAPNSMDAILKWNAP 1422

RESULT 15
AAW24787
ID   AAW24787 standard; protein; 1732 AA.
XX
AC   AAW24787;
XX
DT   17-OCT-2003  (revised)
DT   25-NOV-1997  (first entry)
XX
DE   PrtK antigenic protein complex.
XX
KW   Periodontal disease; cell surface protein; thiol protease; endopeptidase;
KW   PrtK; PrtK48; PrtK39; PrtK15; PrtK44; haemagglutinin; adhesin; therapy;
KW   diagnosis; vaccine; antigen.
XX
OS   Porphyromonas gingivalis; strain W50.
XX
FH   Key              Location/Qualifiers
FT   Peptide          1..228
FT                    /label= Pro-pro_peptide
FT   Cleavage-site    228..229
FT   Protein          229..737
FT                    /label= PrtK48
FT                    /note="48 kDa Lys-specific thiol protease"
FT   Cleavage-site    737..738
FT   Protein          738..1156
FT                    /label= PrtK39
FT                    /note="39 kDa adhesin"
FT   Cleavage-site    1156..1157
FT   Protein          1157..1291
FT                    /label= PrtK15
FT                    /note="15 kDa adhesin"
FT   Cleavage-site    1291..1292
FT   Protein          1292..1732
FT                    /label= PrtK44
FT                    /note="44 kDa adhesin"
XX
PN   WO9716542-A1.
XX
PD   09-MAY-1997.
XX
PF   30-OCT-1996;  96WO-AU000673.
XX
PR   30-OCT-1995;  95AU-00006275.
XX
PA   (UYME ) UNIV MELBOURNE.
PA   (VICT-) VICTORIAN DAIRY IND AUTHORITY.
XX
PI   Reynolds EC,  Bhogal PS,  Slakeski N;
XX
DR   WPI; 1997-272112/24.
DR   N-PSDB; AAT78851.
XX
PT   New antigenic protein complex from Porphyromonas gingivalis - comprising
PT   Arg- and Lys- specific thiol endo-peptidase(s), used in the detection,
PT   prevention and treatment of periodontal disease.
XX
PS   Example 1; Fig 9b; 68pp; English.
XX
CC   A PrtR-PrtK cell surface protein of Porphyromonas ginivalis (PG)
CC   comprises a 300 kDa complex composed a 48 kDa lysine-specific thiol
CC   protease and 39, 15 and 44 kDa adhesins encoded by the prtK gene
CC   (AAT78851), and a 45 kDa arginine-specific thiol protease and 44, 15, 17
CC   and 27 kDa adhesins (see AAW24786) encoded by the prtR gene (AAT78850). A
CC   claimed antigenic complex comprises at least one multimeric protein
CC   complex of PrtR and PrtK each containing at least one adhesin domain, the
CC   complex having a mol.wt. of over 200 kDa, and preferably comprises all 9
CC   proteins of the PrtR-PrtK complex (see also AAW24780-85). It can be used
```

```
CC   in a claimed composition to elicit an immune response directed against
CC   PG, and in a claimed method of reducing the prospect of PG infection
CC   and/or severity of disease. Antibodies directed against the complex are
CC   claimed for use in treating PG cells or other
CC   previously prepared antigens based on fimbriae or the capsule, the PrtR-
CC   PrtK complex or component parts are safe and effective antigens. (Updated
XX   on 17-OCT-2003 to standardise OS field)
SQ   Sequence 1732 AA;

Query Match             30.8%;  Score 812.5;  DB 2;  Length 1732;
Best Local Similarity   30.0%;  Pred. No. 6.4e-53;
Matches 222;  Conservative  70;  Mismatches 172;  Indels 277;  Gaps  22;

Qy     22  TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD--------------PIPAGMARIILEA 67
               ||| :||     |      || :  |        |           :| ||||:
Db    693  TATTQGQKVTLKWEAPS-----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy     68  HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
               ||||||||  :|: ||  :    | ||        |:  ||  |  :|| |:
Db    748  DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVT 806

Qy    127  TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
               ||||||   : ||||:||||:|||||  ||:|:||        |||:| :||| |:|||
Db    807  QNIIVTGQGEVVIPCGVYDYCITNPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFTM 866

Qy    182  QRQGPGDAASVV--------------------------------------------- 193
               :|  ||  :|:
Db    867  RRAGMGDGTDMEVEDDSPASYTYVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy    194  -----------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP---------------ASDKRTYVLN 234
                                | ||||:|| :|||||||| ||||| |||
Db    927  AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPGTTLS 986

Qy    235  ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
               ||| :|| ||| |||||||:|  : |:|| || || ||    |  |    || |  |:||
Db    987  ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042

Qy    295  NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS--- 350
               ||| : ||:|  || ||  ||:|   ||  :| |  |  || ||    |||| |:
Db   1043  NYLITPALDDIPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102

Qy    351  ---------------DKPAPMNLV------------------------------------------KSEGVKL---- 366
                              || |||||                                       |: :
Db   1103  RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy    367  --------PAPYQERTID--------------------------LSAYAGQQV-- 385
                       || | |||                             |:||||||
Db   1163  FESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy    386  --------------------------------------------- 385
Db   1223  LISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFG 1282

Qy    386  ---------------------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
                                                 |:| | | |||| || |  || |||:|| ||
Db   1283  LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNVILLDDIQFTMGGSPTPTDYT 1342

Qy    418  YTVVRDNVIAQNLAATTNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
               ||| :|||:||||||| |||  |||||:||||||||||||||| |||||||||| |:||
Db   1343  YTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTV-NSTQFNPV 1401

Qy    478  QNLTGSAV--GQKVTLKWDAP 496
               ||||:|| ||:||
Db   1402  QNLTAEQAPNSMDAILKWNAP 1422

Search completed: May 18, 2004, 11:42:39
Job time : 56.2572 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:37:00 ; Search time 12.9765 Seconds
(without alignments)
3684.135 Million cell updates/sec

Title: US-08-570-311-2
Perfect score: 2641
Sequence: 1 MRKLNSLFSLAVLLSLLCWG..........QNLTGSAVGQKVTLKWDAPN 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2641	100.0	2628	2	T28651	hemagglutinin A -
2	950.5	36.0	1526	2	S49763	gingipain R (EC 3.
3	944.5	35.8	1704	2	A55426	gingipain R (EC 3.
4	812.5	30.8	1732	2	T30836	lysine-specific cy
5	494.5	18.7	991	2	I40229	arginyl endopeptid
6	156	5.9	1231	1	A48490	endo-1,4-beta-xyla
7	153	5.8	5291	2	F90696	hypothetical prote
8	152	5.8	1234	2	S72640	endo-1,4-beta-xyla
9	150	5.7	1483	2	C97012	probably celluloso
10	147	5.6	1983	2	G86643	hypothetical prote
11	144.5	5.5	781	2	T36143	probable secreted
12	138.5	5.2	2468	2	A83412	hypothetical prote
13	137.5	5.2	635	2	S19011	endo-1,4-beta-xyla
14	137	5.2	2013	2	AI1489	probable peptidogl
15	135.5	5.1	1433	1	A36734	bacillopeptidase F
16	135	5.1	1655	2	B97835	hypothetical prote
17	132	5.0	1651	2	JC1340	outer membrane pro
18	131.5	5.0	725	2	A90255	hypothetical prote
19	129.5	4.9	1873	2	T30944	surface protein pr
20	129	4.9	940	2	ABI1744	internalin protein
21	129	4.9	1349	2	AI1476	cell surface prote
22	128	4.8	1348	2	AH1115	cell surface prote
23	128	4.8	4936	2	AH2515	hypothetical prote
24	127.5	4.8	5188	2	B85547	probable RTX famil
25	127	4.8	2710	2	A37052	toxin A - Clostrid
26	126.5	4.8	908	2	AE2254	hypothetical prote
27	126.5	4.8	1148	2	S72635	exo-poly-alpha-gal
28	125.5	4.8	940	2	AD1374	internalin protein
29	125	4.7	1503	2	T18266	cycloinulo-oligosa
30	124.5	4.7	868	2	AF3204	autotransporter pr
31	124.5	4.7	875	2	H90371	proteinase limport
32	124	4.7	1616	2	T17884	S-layer protein -
33	124	4.7	2013	2	AD1129	probable peptidogl
34	123.5	4.7	208	2	AF1718	hypothetical prote
35	122	4.6	729	2	T35028	probable glycosyl
36	122	4.6	1217	2	F97177	alpha-glucosidase
37	121.5	4.6	1530	2	AH1396	peptidoglycan anch
38	121	4.6	1142	2	C97080	levanase/invertase
39	120.5	4.6	587	2	A55368	transferred entry
40	120.5	4.6	1448	2	AI2007	Subtilase family p
41	120	4.5	364	2	JE0292	fungal stress prot
42	120	4.5	1939	1	D97316	probable S-layer p
43	119	4.5	699	2	A38368	chitinase (EC 3.2.
44	119	4.5	2205	2	T08615	aggregation factor
45	118.5	4.5	694	2	B84331	hypothetical prote

ALIGNMENTS

RESULT 1
T28651
hemagglutinin A - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C;Accession: T28651
R;Han, N.; Whitlock, J.; Progulske-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A;Title: The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381 contains four :
A;Reference number: Z20494; MUID:97047672; PMID:8926061
A;Accession: T28651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2628 <HAN>
A;Cross-references: EMBL:U41807; NID:g1552410; PID:g1469916; PIDN:AAB17128.1
C;Genetics:
A;Gene: hagA

Query Match 100.0%; Score 2641; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 2.1e-163;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60

Qy   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120

Qy  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIIYVGEGVSKGNDYVVEAGKTYHFT  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIIYVGEGVSKGNDYVVEAGKTYHFT  180

Qy  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240

Qy  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300

Qy  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKRPAPMNLVK  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKRPAPMNLVK  360

Qy  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYYTYV  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYYTYV  420

Qy  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
```

```
Db   421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL 480
```

```
Db   416 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA 475
             ||||||||||||||||||||||||||||||||||||||||||||||| .:||
Qy  1323 YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN 1381
```

```
Qy   481 TGSAVGQKVTLKWDAPN 497
             |||||||||||||||||
Db   481 TGSAVGQKVTLKWDAPN 497
```

```
Qy   476 HVQNLTGSAVGQKVTLKWDAPN 497
             :|||
Db  1382 PVKNLKAQPDGGDVVLKWEAPS 1403
```

RESULT 2
S49763
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis (fragment)
C;Species: Porphyromonas gingivalis
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997
R;Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.
submitted to the EMBL Data Library, November 1994
A;Description: Cloning, sequence analysis and expression in Escherichia coli of prpRl of
A;Reference number: S49763
A;Accession: S49763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <ADU>
A;Cross-references: EMBL:X82680
C;Genetics:
A;Gene: prpR1
C;Keywords: cysteine proteinase; hydrolase

```
Query Match     36.0%;  Score 950.5;  DB 2;  Length 1526;
Best Local Similarity   31.8%;  Pred. No. 1.1e-53;
Matches 236;  Conservative 77;  Mismatches 154;  Indels 275;  Gaps   20;
```

RESULT 3
A55426
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
N;Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP; :
C;Species: Porphyromonas gingivalis
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
C;Accession: A55426; D53113
R;Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr,
J. Biol. Chem. 270, 1007-1010, 1995
A;Title: Molecular cloning and structural characterization of the Arg-gingipain protein
A;Reference number: A55426; MUID:95138080; PMID:7836351
A;Accession: A55426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1704 <PAV>
A;Cross-references: GB:U15282; NID:g557067; PIDN:AAA69539.1; PID:g557068
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isola
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: D53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 228-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141694)
C;Keywords: cysteine proteinase; hydrolase

```
Query Match     35.8%;  Score 944.5;  DB 2;  Length 1704;
Best Local Similarity   31.9%;  Pred. No. 3.2e-53;
Matches 235;  Conservative 75;  Mismatches 160;  Indels 267;  Gaps   20;
```

```
Qy    22 TAAAQG------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH  68
          :||:|          :|.||.:|        |||  :|.||::|||:|||||||||||
Db   671 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730
```

```
Qy    69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
          ||::||:|||||:|||||||||:||     ||||||||||||:||||||||||||||||||
Db   731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789
```

```
Qy   127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
          ||:|:||||||:|||||||||||||||||  |:|||||:|||||||||:|||||:|||||
Db   790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849
```

```
Qy   186 PGDAASVV------------------------------------------------- 193
          ||||:|
Db   850 SGDGTELTISEGGSGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909
```

```
Qy   194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS----------DKRTYVLNESF 237
                :||||||||||||||||||||||||||||||||         ||||||:|||
Db   910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESF 969
```

```
Qy   238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
          :||   :|||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db   970 B-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025
```

```
Qy   298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS------ 350
          :||  :||||| ||| |||     |||||:|||||||||||||||||||||||||
Db  1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEETITAKGVRSP 1085
```

```
Qy   351 -------------------DKPAPMNLV---------------------KSEGVKL--- 366
                               :|           |||:|||:
Db  1086 EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145
```

```
Qy    22  TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARILEA     67
Db   693  TATTQQKVTLKWEAPS------AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA  747
Qy    68  HDVWEDGTGYQMLWIDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
Db   748  DNVWGDNTGYQFLLDADHNTPGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVTT  806
Qy   127  TNFVLDGTASADIPAGTYDVYIINPNP--GIIYIVEG---VSKGNDYVVEAGKTYHFTV   181
Db   807  QNIIVTGQGEVVIPGGVYDYCITNEPRASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM  866
Qy   182  QRQGPGDAASVV-----------                                      193
Db   867  RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT  926
Qy   194  -------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP-----ASDKRTYVLN        234
Db   927  AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS    986
Qy   235  ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD  294
Db   987  ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042
Qy   295  NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS-    350
Db  1043  NYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102
Qy   351  ----------------------DKPAPMNLV----KSEGVKL                   366
Db  1103  RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162
Qy   367  --PAPYQERTID---------------LSAYAGGQV                         385
Db  1163  FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGSNVVSSFSWNGMALNPDNY  1222
Qy   386  ---------------                                              385
Db  1223  LISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEETPNGINKGGARFG 1282
Qy   386  -----------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT                417
Db  1283  LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT 1342
Qy   418  YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV  477
Db  1343  YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTV-NSTQFNPV 1401
Qy   478  QNLTGSAV--GQKVTLKWDAP                                        496
Db  1402  QNLTAEQAPNSMDALLKWNAP                                       1422
```

```
Qy   367  ---PAPYQERTID---------------LSAYAGGQV--                      385
Db  1146  STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205
Qy   386  ---------------                                              386
Db  1206  KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST 1265
Qy   386  -----------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV             420
Db  1266  EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQPTMGGSPTPTDYTYTV 1325
Qy   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
Db  1326  YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1384
Qy   481  TGSAVGQKVTLKWDAPN                                            497
Db  1385  KAQPDGGDVVLKWEAPS                                           1401
```

RESULT 4

T30836

lysine-specific cysteine proteinase porphypain (EC 3.4.22.-) - Porphyromonas gingivalis
N;Alternate names: lysine-specific cysteine proteinase 1, 60K
C;Species: Porphyromonas gingivalis
C;Date: 22-Oct-1999 #sequence revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30836; T30837; T30526; A53113
R;Barkocy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulske-Fox, A.; Lantz,
J. Bacteriol. 178, 2734-2741, 1996
A;Title: Analysis of the prtP gene encoding porphypain, a cysteine proteinase of Porphyr
A;Reference number: Z20895; MUID:96213011; PMID:8631659
A;Accession: T30836
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 11732 <BAR>
A;Cross-references: EMBL:U42210; NID:g1314325; PID:g1314326; PIDN:AAB06565.1
R;Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z20896
A;Accession: T30837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-795,'I',797-1389,'N',1391-1478,'Y',1480-1732 <SLA>
A;Cross-references: EMBL:U75366; NID:g2182811; PID:g2182812; PIDN:AAB60809.1
R;Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A;Title: IS195, an insertion sequence-like element associated with protease genes in Por
A;Reference number: Z20844; MUID:98298016; PMID:9632563
A;Accession: T30526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1350,'N',1352-1363,'Y',1365-1447,'H',1449-1732 <LEW>
A;Cross-references: EMBL:AF017059; NID:g2738802; PID:g2738803; PIDN:AAC26523.1
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: A53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 229-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141690)
C;Genetics:
A;Gene: prtP; prtK
C;Keywords: cysteine proteinase; hydrolase

Query Match   30.8%; Score 812.5; DB 2; Length 1732;
Best Local Similarity 30.0%; Pred. No. 1.3e-44;
Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;

RESULT 5

I40229

arginyl endopeptidase - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40229
R;Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A;Title: Structural characterization of arginipain, a novel arginine-specific cysteine
A;Reference number: I40229; MUID:95168894; PMID:7864651
A;Accession: I40229
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-991 <RES>
A;Cross-references: GB:D26470; NID:g927644; NID:g927645; PIDN:BAA05484.1; PID:g927645

Query Match   18.7%; Score 494.5; DB 2; Length 991;
Best Local Similarity 25.5%; Pred. No. 2.9e-24;
Matches 124; Conservative 41; Mismatches 90; Indels 231; Gaps 7;

RESULT 6
A48490
endo-1,4-beta-xylanase (EC 3.2.1.8) A precursor - Thermoanaerobacterium saccharolyticum
C;Species: Thermoanaerobacterium saccharolyticum
C;Date: 03-May-1994 #sequence_revision 11-Apr-1997 #text_change 16-Jul-1999
C;Accession: A48490
R;Lee, Y.E.; Lowe, S.E.; Henrissat, B.; Zeikus, J.G.
J. Bacteriol. 175, 5890-5898, 1993
A;Title: Characterization of the active site and thermostability regions of endoxylanase
A;Reference number: A48490; MUID:93388520; PMID:8376336
A;Accession: A48490
A;Molecule type: DNA
A;Residues: 1-1231 <LEE>
A;Cross-references: GB:M97882; NID:g533366; PIDN:AAA21812.1; PID:g533367
A;Experimental source: strain B6A-RI
A;Note: mutation experiments support roles for Asp-537, Glu-600, and Asp-602; His-572 ca
-495, Asn-539, and His-572 are supported by X-ray reference A55905 on another molecule w
C;Genetics:
A;Gene: xynA
C;Function:
A;Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans
A;Pathway: xylan degradation
C;Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;
a xylanase A cellulose-binding repeat homology
C;Keywords: duplication; glycosidase; heat-stable protein; hydrolase; polysaccharide deg
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-1231/Product: endo-1,4-beta-xylanase A #status predicted <MAT>
F;39-186/Domain: endo-1,4-beta-xylanase A amino-terminal repeat homology <TXA1>
F;196-340/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA2>
F;386-676/Domain: Thermotoga xylanase A homology <SXY>
F;683-854/Domain: Streptomyces endo-1,4-beta-xylanase A homology <TXC1>
F;858-1043/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC2>
F;1056-1109/Domain: S-layer repeat homology <SLR1>
F;1115-1168/Domain: S-layer repeat homology <SLR2>
F;1179-1231/Domain: S-layer repeat homology atypical <SLR3>
F;495,572,600,602/Active site: Glu, His, Glu, Asp #status predicted

```
Qy   22 TAAAQG-------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARILEAH  68
           ||||         :||         :  :  |||         :: |||||||
Db  671 TATTQGQKVTLKWDAPSTKNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
           |||:||:||| ||||| ||| :||:          ||:||||| |||:|||||| ||
Db  731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPEYTVPENADPSCSP  789

Qy  127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
           ||::|||||::||||||||::|:: :|| ||| |:|:  |:|||:||||:|||:: : |
Db  790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 849

Qy  186 PGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTVLNESFDTQTLPNG  245
           ::
Db  850 SGDGTELTIS-----                                              859

Qy  246 WTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVP 305
Db  860 -----                                                       859

Qy  306 ENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVKSEGVK 365
Db  860 -----                                                       859

Qy  366 LPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTVYRDNV 425
                    ||||||||||||||||||
Db  860 -----           EGGGSDYTYTVYRDGT                            875

Qy  426 VIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAV 485
                :: ::|||||||||||||
Db  876 KIKEGLTETTYRDAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVV 935

Qy  486 GQKVTL 491
             :|::|
Db  936 GKTITV 941
```

F;537,539/Binding site: substrate (Asp, Asn) #status predicted

Query Match 5.9%; Score 156; DB 1; Length 1231;
Best Local Similarity 20.4%; Pred. No. 0.043;
Matches 92; Conservative 64; Mismatches 128; Indels 166; Gaps 23;

```
Qy  106 AGLYDPFEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSK 165
Db   22 ASLFSP---------PIRVFADDTNINLVSNG---DFESSTIDGWIKQGNPTLAVTTEQAIGQ   72

Qy  166 GNDYVVEAGKTYHFTVQRQP-------GDAASV-----VVTGS-GGNEFAPVQNLQ     209
Db   73 YSMKVTGRTQTY------EGPAYSFLGKMQKGBSYSVSLKVRLVSGQNSSNPLITVTMFR   126

Qy  210 WSVSGQTV-TLTWQAPASD--------------KRTVLNESFD                    238
Db  127 EDDNGKHYDTIVWQKQVSEDSWTTVSGTYTLDYIGTLKTLYMYVESPDPTLEYYIDDVVV  186

Qy  239 -----------------TQTLPNGWTMIDADGDGHNWLSTIN---VYNTATHTGDGAMFSKS 280
Db  187 -------------------------------SGWIGTGSSVVKAVYGVA-HSGDYSLLITG 237

Qy  281 WTAS-GGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANF 339
Db  238 RTANWNGPSYDL-------------------TGKIVPGGQYNVDFWVK----------FVNGNDTEQIKA 278

Qy  340 TIKLLEETLGSDKPAPM-----------NLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFN 393
Db  279 TVKAT------SDKDNYIQVNDFANVNKGEWTEIKGSF---TLPVADYSGISIIVESQ-N 328

Qy  394 STGIFRLYLDDVAVSGEGSSNDYT----------YTVYRD                       423
Db  329 PT-LEFYIDDFSVIGEISNNQITIQNDIPDLYSVFKDYFPIGAVDPSRLNDADPHAQL  386

Qy  424 ------NVVIAQNLAATTFNQENVAPGQYNY                                448
Db  387 TAKHFNMLVAEN----AMKPESLQPTEGNF                                412
```

RESULT 7
F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain R
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: F90696
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90696
A;Molecule type: DNA
A;Status: preliminary
A;Residues: 1-5291 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33965.1; PID:g13360000; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0542

Query Match 5.8%; Score 153; DB 2; Length 5291;
Best Local Similarity 24.3%; Pred. No. 0.52;
Matches 137; Conservative 70; Mismatches 202; Indels 154; Gaps 33;

```
Qy     5 NSLFSLAVLLSLLCWGQTAAAQG--GPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGMA-  61
Db  3303 NNGYLTIATVSDLAGNLGSASKGVTVDTTAPVISNTVAGDDVINNVEHIQAQIISGTAT  3362

Qy    62 ------RIILEAHDVWEDGTGYQMLWDADHNQYGASIPESF-WFANG-TIPAGLYDPF  112
Db  3363 GAVAGDRLV-----VTIAGQYVITSDASGN-WSVGVPASVISGLADGTVTISATITD- 3414

Qy   113 EYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVE 172
```

```
Db          :|| :||| |:|  |||   ||| :                          :.||
28   IRAFADDTNTNLVSNG-----DFETGTIDGWIKQGNPTLEVTTEQAIGQYSMKVTGRTQTY   83

Qy          :||                          .:|   ||                   ::|::||
173  AGKTYHFTVQRQPGDAASVVVT-----GEGGNEFAPVQ-NLQWSVSGQTVTLTWQAPASD    227
Db   178  HFTVQRQGP----------GDAASV------VVTGE-GGNEFAPVQNLQWSVSGQTV-TLT   220

Qy   3451 AGSA--LILSGTGTNFATGTVVTVLLNGKGYS--ATIQSNGSWSVN---VPAADVAALSD   3503
Db    84  -----EGPAYSFLGKMQKGESYNVSLKVRLVSGQNSSNPFITVTMFREDDNGKHYDTIV    137

Qy   228  KRTVLNESFDTQTLPNGWTMIDADGDGHNWLSTINV-YNTA-------THTGDGAMFSK    279
Db   221  WQAPASDKRTYVLNESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKS   280

Qy   3504 GTSYTVSASAQ-----------DSAGNGNSSTQTHNVQVNTAAVSLSVSTISGDNLINA-   3551
Db   238  WQKQVSE-----------------DSWTTVSGTYTLDYTGT--LKTLYMY            168

Qy   280  SWTASGGAKIDLS--PDNYLVTPKVTVPENGK-LSYWVSSQVPWT----------NE    323
Db   281  WTASGGAKIDLSPDNYLVTPKVTVP--------ENGKLSYWVSSQVPWTNEHYGV---   327

Qy   3552 --AEAGSALITLSGTGTNFATGTVVTVLLNGKGYSATIQSNGSWSVNVPAADVAALSDGT   3608
Db   169  -VESPDPTLEYYIDDVVVTPQNPIQVGNVITNGTFENGNTSGWVGTGSSVVKAVVGVAHS   227

Qy   324  HYGVFLS--TTGNEAANFTIKLLEETLGSDKPAP--MNLVKSEGVKLPAPYQERTIDL    377
Db   328  ---FLSTTGNEA-----------------ANFTIKLL------EETLGSDKPAP     355

Qy   3609 SYTVSASAQDSAGNSAT------ASRSVAVDLTAPVISINTVSTDD-RLNAAEQQQPLTL   3661
Db   228  GGYSLLTTGRTANWNGPSYDLTGKIVPGQQYNVDFWVKFVNGNDTEQIKATVKATSNKDN   287

Qy   378  ---SAYAGQQV----------YLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTVYRDNV   425
Db   356  ------MNLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVS   408

Qy   3662 NGSTSAEVGGTVTVTFGGKTYTATVAANGTWALNVPAVDLAALGQG----------   3707
Db   288  TLPVTDYSGVSIYVESQ--NPT--LEFYIDDFSVI     340

Qy   426  VIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTG---   482
Db   409  GEGSSNDYT-----------YTVYRD                              433

Qy   3708 -AQTTTASV-NDRAGNPGQATHALTVD---TVAPTV-TIATVAGDDIINNAEQLAGQTI   3760
Db   341  GEISNNQITIQNDIPDLYSVFKDYFPIGVAVDSSRLNDADPHAQLTAKHFNMLVAEN---   397

Qy   483  --SAVGGKVTL-----KWDA    495
Db   434  TTFNQENVAPGQYNYCVE-----VKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQK   488

Qy   3761 SGTTTAEVGQTVTVTFNGQTWSA   3783
Db   398  -AMKPESLQPTEGNFTFDNADKIVDY-----------------EIAHNMOMRGH----   433

Qy   489  VTLKW 493
Db   434  -TLLW 437
```

RESULT 8
S72640
endo-1,4-beta-xylanase (EC 3.2.1.8) xynA precursor - Thermoanaerobacterium thermosulfurigenes
C;Species: Thermoanaerobacterium thermosulfurigenes
C;Date: 29-Jul-1997 #sequence_revision 29-Jul-1997 #text_change 16-Jul-1999
C;Accession: S72640; S72622
R;Matuschek, M.; Sahm, K.; Bahl, H.
A;Description: Characterization of genes from Thermoanaerobacterium thermosulfurigenes E
submitted to the EMBL Data Library, March 1996
A;Reference number: S72640
A;Accession: S72640
A;Molecule type: DNA
A;Residues: 1-1234 <MATU>
A;Cross-references: EMBL:U50952; NID:g1255235; PIDN:AAB08046.1; PID:g1255238
A;Experimental source: strain EM1
R;Matuschek, M.; Sahm, K.; Zibat, A.; Bahl, H.
A;Title: Characterization of genes from Thermoanaerobacterium thermosulfurigenes EM1 tha
Mol. Gen. Genet. 252, 493-496, 1996
A;Reference number: S72621; MUID:97033555; PMID:8879252
A;Accession: S72622
A;Molecule type: DNA
A;Residues: 815-1234 <MAW>
A;Cross-references: EMBL:U50952
C;Genetics:
A;Gene: xynA
C;Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;
a xylanase A cellulose-binding repeat homology
C;Keywords: glycosidase; hydrolase
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-1234/Product: endo-1,4-beta-xylanase XynA #status predicted <MAT>
F;38-185/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA1>
F;195-339/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA2>
F;385-675/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
F;682-853/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC1>
F;857-1042/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC2>
F;1055-1108/Domain: S-layer repeat homology <SLR>

```
Query Match          5.8%;    Score 152;   DB 2;   Length 1234;
Best Local Similarity   20.0%;   Pred. No. 0.078;
Matches   97;   Conservative   66;   Mismatches   138;   Indels   184;   Gaps   23;
```

```
Qy   118  VNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTY    177
```

RESULT 9
C97012
probably cellulosomal scaffolding protein precursor, secreted, cellulose-binding and col
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: C97012
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C97012
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1483 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78886.1; PID:g15023809; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0910

```
Query Match          5.7%;    Score 150;   DB 2;   Length 1483;
Best Local Similarity   20.2%;   Pred. No. 0.14;
Matches  127;   Conservative   81;   Mismatches   213;   Indels   208;   Gaps   29;
```

```
Qy   27   GGPKTAP----------SVTHQAVQKGIRTSKVKDLRDPIPAGMARIILEAHDVWEDGTGY-   77
Db   522  GTPVVTPSQINVEQGSATDQPVKIDLNGNTLKDVVDQ---SGKTLVQGTDYTVTDGIT    577

Qy   78   ----------QMLWDADHNQYGAS----------IPEESFWFANGTIPAGLYDPFEYKVPV   118
Db   578  LSQSYLAGLALGQYTLTLDFNGGGASQTITINVKNETVKLSVGTVSGNPGD--TVKVPV    635

Qy   119  N----------ADASFSPTNF-VLDGTASADIPAGT--YDYVIINPNPGIIYIV-----   159
Db   636  TISQVSTPVGLICMDISYDASKFTVKDVLPNTDLVKDTDNYSFIVNTSTPGKISITFTDP   695

Qy   160  ----------GEGVSKGNDYVVEAGKT--YHFTVQRQPGDAASVVVTGEGGNEFAPVQNL   208
```

RESULT 10
G86643
hypothetical protein ybeF [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: G86643
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: G86643
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1983 <STO>
A;Cross-references: GB:AE005176; PID:g12723000; PIDN:AAK04249.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: ybeF

Query Match   5.6%; Score 147; DB 2; Length 1983;
Best Local Similarity 21.7%; Pred. No. 0.32;
Matches 117; Conservative 77; Mismatches 181; Indels 164; Gaps 30;

```
Qy   39 AVQKGIRTSKVKDLRDPIPAGMARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFW 98
         || :| :|| |||| || ||| |
Db  628 ATQQGI---VKNIDGDI------------DYENREIPWKIDINSPG-------YW 660

Qy   99 FANGTIPAGLYDPFEYKVPVNADASFSPTNF-VLDGTASADIPAGTYDVIINPNPGIIY 157
         | :| :|                       :| ||
Db  661 MENWSL---------EDKMSEGLTFLENTFQIIDKTAGNKVLSPT-EYTLIKTTAGFSV 709

Qy  158 IVGEGVSKGND--YVVEAGKTYHFTVQRQPGG--------DAASVVVTGEGGN---- 200
         | : | | |    :|
Db  710 SFNSPLKEGTDHKYQIKYKTKFDTSVIDNGSGHEGDIKFVNDASMTWKDKNGGDHTNNDH 769

Qy  201 -EFAPVQNLQWSVSGQTVLTWQAPASDKRTYVLNESFDTQTLPNGWTMIDADGDGHNWL 259
          : :|                            ||
Db  770 KEFKPIPPPQY--NGQK-SGSYNA-TSKKITWTIAANFNQQELSNA-SITDPISDDQNYV 824

Qy  260 S---------TINVYNT----ATHTGDGAMFSKSWTASGSGAKIDLSPDN----YLVTPKVT 303
         |          |: |||     | ||||||||||||||||
Db   ... TINVYNT----ATHTGDGAMFSKSWTASGSGAKIDLSPDN----YLVTPKVT
```

```
Db  825 SGSAKVYEATINKNGTYTLGABVTSD--MGIKIVESKGSVKVEL-PDGSTKAYVLIFETS 881
         || :| :||
Qy  304 VP---------------------ENGKLSYWVSSQVPWTNEHYGVFLSTTGNE- 335

Db  882 LEGNLINQKEYXNKATFTNKDISHDLSASV--TPAHQGEFVTKDGSQSSTDSNYVNWKLT 939
         : :|: ||         |||
Qy  336 --AANFTIKLLEETLGSDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFN 393

Db  940 VNASQSTLKNVEVT--DNPSSNQYIVAKDILI------YGTSIDASG-N 979
         |: || |            | ::
Qy  394 STGIFRLYLDDVAVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVK 453

Db  980 ITENKNIILE----------QGKDYSVDIQTDNSTGAQTVKIKFLSEINTA---YVVE- 1024
           |:|| :
Qy  454 YTAGVSPKVCKDVT-----VEGSNE----FAHVQNLTGSAVGQK--VTLKWNDAP 496

Db 1025 YRALIITSDKANDVVTNQAHITGDNEKIIEQDVEKDVPVTNHNGSANGSKGSVTLEKVGP 1083
         |||:| :||
Qy    ... YRALIITSDKAND... VTLEKVGP
```

RESULT 11
T36143
probable secreted proteinase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2003
C;Accession: T36143
R;Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A;Reference number: 221598
A;Accession: T36143
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-781 <SEE>
A;Cross-references: EMBL:AL096852; PIDN:CAB51001.1; GSPDB:GN00070; SCOEDB:SCE19A.20c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE19A.20c
C;Superfamily: secreted virulence protease, InhA type

Query Match   5.5%; Score 144.5; DB 2; Length 781;
Best Local Similarity 21.2%; Pred. No. 0.13;
Matches 111; Conservative 58; Mismatches 164; Indels 191; Gaps 31;

```
Qy   24 AAQGGPKTAP-SVTHQAVQKGIRTSKVKDLRDPIPAG---MARIILEAHDVWE----DG 74
         |: |            |:|       | |::
Db  205 ARYGSNKCDPDNCAWYAVQDGV-TAWIADQK--AAGRTDAQIKSQLTQYDQWDRYDFPG 260

Qy   75 TG---------------YQMLWDADHNQYGASIPEESF---WFANGTIPAGLYDPFEYKVPV 118
         :|
Db  261 DGDFNEPDGYIDHFQIVHAGEDESAGGAQGEDAIWAHRWYAFGT-DAGATGPADNK--- 316

Qy  119 NADASFSPTNFVLDGTASADIPAGTYDVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTY 177
         :| ::
Db  317 ----------------LGGTQIGDTGIWGDYTIQPENGLGVY----AHEYGHDLGLPD 356

Qy  178 HFTVQRQPGDAASVVVTGEGGNEPAPVQNLQWSV--SGQTVLTWQAPASDKRTYVLNE 235
          : :|
Db  357 HY---------DTAG---GDNSTGF--------WTLMSSG---SWLGTGR-----NE 385

Qy  236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASG- 285
         :|   |||  |||||
Db  386 IGDLPGDMNAWDKLQL----GWLN----YDTA----KAGVNSWHKLGLAEVNTKHKQ 430

Qy  286 GAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTI--- 341
         || |:
Db  431 GLVVELPKEK--VTTEIVTPAEGETQWWSGSGNDLKNT-LSRSVDLTGKSAASLTLDGWY 487

Qy  342 ----------------KLLEETL----------GSDKPAPMNLVKSEGVKLPAPYQE 372
                            ||  |:         ||||||||||
Db  488 DIEADYDFLYTEVSTDGGANWTAIDGTFPDGNPIQRDGSDKPALLSATVDAYG-KLVYP--- 543

Qy  373 RTIDLSAYAGQQVYLAFRHFNSTGIFR--LYLDDVAVSGEGSSNDYTYTVYRDNV--- 425
         : |:|||||                    |||| ||
Db  544 ----LDAYAGKKIDLRFRYQTDGGLAMKGFTADEIAVTADGE------TLFSDNAETADD 593
```

Qy   426 ------------------------------VIAQNLAATTFNQENVAPGQYNY 448
                                          :: :   ::: :||||
Db   594 AWTAVGFTRKGASFTKEYAQYYIAENRQVVSYD-KTLKTGPYNF 636

RESULT 12
A83412
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83412
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83412
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2468 <STO>
A;Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1874

Query Match   5.2%;   Score 138.5; DB 2; Length 2468;
Best Local Similarity   23.7%;   Pred. No. 1.6;
Matches 107; Conservative 47; Mismatches 166; Indels 131; Gaps 23;

Qy   103 TIPAGLYDFPEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIYIVGEG 162
           ::| ||| | | :::::|:|||||||||||||||||||||
Db   319 TDPAG------NNSTPVTVEAP----------DTTA---PAPATD--------VQVAPDG 351

Qy   163 VS-KGNDYVVEAGKTYHFTVQRQPGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTW 221
           |: : ||||| || || :|| ||| : |||||||||| :||||||| ||||||
Db   352 SSVTGN--AEPGAT--VGVTDGDGQPDTTVVVGPGGSFEVPLN--PPLTNGETVTVIV 404

Qy   222 QAPASDKRTVLNESFDTQTLPN---------GWTMIDADGDGH---------- 256
           ||: | ::: || | || ||:||:|||||
Db   405 TDPAGNSSTPVTAEAPDFPDAPQVNASNGSVLSGTAEAGVTIVITDGNGNPIGQTSADAN 464

Qy   257 -NWLST--------INVY------NTATH-TGDGAM------------FSKSWTASG 285
            :||:| ::| ||:|| |||:| ::|:|| | |
Db   465 GNWSFTPGSQLPDGTVVNVARDAAGNSSPAISITVDGVAPNAPVVEPSNGSELSGTAEP 524

Qy   286 GAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLE 345
           ||: ||| ||| |||| ||| ||| |:|: : |||||:||||||| ||: |:||| ||
Db   525 GSSVTLTDGNGNPIGQTTADANGNWSFTPSTPLP-----DGTVVNVVARDAAGNSSPPAS 579

Qy   346 ETLGSDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDV 405
           | || : : :|::::| :| || ||||||:|||||||||||||| :: ::|| |
Db   580 VTVDAVAPATPTVDPSNGT------TLSGTAEPGSSVTLTDGNGNPIG----------- 621

Qy   406 AVSGEGSSNDYTYT---VYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPK 461
           ||: |||:||||:   :|||||| :||||||||||||||||||||||||||||||||||
Db   622 QVTADGSGN-WTFTPSTPLPNGTVV--NATATDPSGNASSPAS----VTVDAVAPATPV 673

Qy   462 V--CKDVTVEGSNEFAHVQNLT--GSAVGQ 487
           | : :|:||||:|| ||||:| |||||
Db   674 VNPSNGCTTLSCTAEPGATVTLTDGNGNPIGQ 704

RESULT 13
S19011
endo-1,4-beta-xylanase (EC 3.2.1.8) - Bacillus polymyxa
C;Species: Bacillus polymyxa
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 15-Oct-1999
C;Accession: S19011
R;Gosalbes, M.J.; Perez-Gonzalez, J.A.; Gonzalez, R.; Navarro, A.
J. Bacteriol 173, 7705-7710, 1991
A;Title: Two beta-glycanase genes are clustered in Bacillus polymyxa: molecular cloning,

A;Reference number: S19011; MUID:92041687; PMID:1938968
A;Accession: S19011
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-635 <GOS>
A;Cross-references: EMBL:X57094; NID:g48815; PIDN:CAA40378.1; PID:g48816
A;Note: the authors translated the codon GAA for residue 78 as Gly, CCT for residue 272
C;Function:
A;Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans
A;Pathway: xylan degradation
C;Superfamily: Clostridium xylanase A repeat homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;408-502/Domain: Clostridium xylanase A repeat homology <CXA>

Query Match   5.2%;   Score 137.5; DB 2; Length 635;
Best Local Similarity   21.2%;   Pred. No. 0.27;
Matches 123; Conservative 65; Mismatches 176; Indels 217; Gaps 32;

Qy   22 TAAAQGGPKTAP---SVTHQAVQKGIRTSKVKDLRDPIPAGMARIILEAHDVWEDGTGY 77
           ||| : | :||: | |
Db   156 TADTPIGPWTDPLGKALVTHST---------PGMAGVTWLFDPAVLVDDDGTGY 200

Qy   78 QMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKV-----------PVNADA 122
           |:: :   : ||: :: ::||
Db   201 --LYS-----GGGIPNES---------DPASIANPKTARVIKLGADMTSVIGSATTIDAPY 245

Qy   123 SFSPT----------------NFVLDGTASADIPAGTYDYVII-NP---------NPG 154
           :|| :|| |:: | | :|:
Db   246 LFEDSGIHKYNGKYYYSYCINFA--GTHPQQYPAGEIGYMVSDNPMGPFTYKGHFLKNPY 303

Qy   155 IIYIVG---------------------------EGVSKGNDYYVEAGKTY---HFT-VQRQPGPGDAA 190
             :|: | |:| |: | |::|
Db   304 TFFGVGGNNHHAVFNFKNEWYVVYHAQTVSKAQ---IGAGKGYRSPHINKLVHKEDGSIS 360

Qy   191 SVV--VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQTLPNGWTM 248
           | | :| : ||:||| |:| |:
Db   361 EVQGNMTG------IAQLSMNMPYTRVEAETIAWQAGVTTEPT--------------- 397

Qy   249 IDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAK------------IDLSP 293
           : : |: | :|| :| :: :|
Db   398 -QASGGP--ISNLNVTN--IHNGDWIAVGKADFGSAGAKTFKANVATNVGGNIEVRLDS 451

Qy   294 DNYLVTPKVTVPENGKLSYW--VSSQVPWTNEHYGVFLSTTGNEAAN---------FTIKL 343
                      |||| | :   :: |||||:|||||:
Db   452 ETGPLVGSLKVPSTGGMQTWREVETTINNATGVHNIYLVFTGSGSGNLLNLDAWQFTPNT 511

Qy   344 LEETLGSDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHF--NSTGIFRLY 401
           :: : | : || : :: || :||:| |||
Db   512 GGNTITKVEAEMKIGGTYAGKISAPFD---GVALYANAD-YVSYSQYFANST----- 560

Qy   402 LDDVAVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPK 461
            :   :: ::::   :
Db   561 -HNISVRG-ASSNAGTAKV---DLVIG----------GVTVGSFNF------TGKTPT 597

Qy   462 VCKDVTVEGSNEFAHVQNLTGSAVGQKVTL------KWDA 495
                 ::: || :
Db   598 V---------QTLSNITHATGDQEIKLALTSDDGTWDA 626

RESULT 14
AI1489
probable peptidoglycan bound protein (LPXTG motif) lin0457 [imported] - Listeria innocua
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AI1489
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AI1489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2013 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95689.1; PID:gi16412898; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0457

Query Match   5.2%;   Score 137;   DB 2;   Length 2013;
Best Local Similarity   20.9%;   Pred. No. 1.5;
Matches   103; Conservative   82; Mismatches 163; Indels 144; Gaps   30;

Qy   74 GTGYQM.LWDADHNQYGASIPEES---FWFANGTIPAGLYDPFEYKVPVNADASFSPTNF 129
              ||::| :|||::| | : ||::::|:||||:||::|::|::|| |||
Db   1185 GQDYEIALFDKDGKEVQSAIKTNNKGEFSFSD---VAIKNPADFKLKVTAP---TRTNF 1237

Qy   130 VLDGTASADIP---AGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGK-TYHFTVQRQG 185
              || : || : ||::|::| |||:||:|||| |:| | :|||| ||::|:|:::
Db   1238 VY---SAKNPLFNMSTKEYTLNSVVPGV------GGVAE--IYITETSKPTKIILDKAV 1286

Qy   186 PGDAASVVVTGEGGNEFAPVQNLQWSV----------SGQTV-----------TLTW 221
              | :|| :||| | | ||:|:|| : ||::| |:|
Db   1287 TPNA----ITIESSDRATEVTN-EWTVVDSNGTVVVSGTGNTIRIPNDEGTYIAKNTATD 1341

Qy   222 QA--PASDKRTVLNESFDTQPLPNGWTMIDADGDGH------NWLSTINVYNTATHTGD 273
              :| |||:|||: || ::|| :||:: ::| ||| ||:|:|::|::|::||
Db   1342 EAGNTASDEKTFDIDYTVPTLTVNQ----DASAEVNSTEANIDWIKPLNVAATDTHDGN 1396

Qy   274 --GAMFSK-SW-----------TASGGAK----------IDLSPDNYLVTPKVTVPENG 308
              ::: ::| :| |||: || :|:|||::|:||:|::|:||
Db   1397 ITPVVDYSKVKNDVLGTYPVTVTATDASGNKATQTVNLRIVDTTSPTLIT--------NN 1449

Qy   309 KLSYWVSSQVPWTNEHYGVFLSTTGNE---------------AANFTKLLEETLG 349
              |:|:|:||:|||:|:||:|::|:||:   :||::|::|::||
Db   1450 PLTYSIENMRKLTEQELYKAAGLIGGDNYDLAPGQSVQPNKQPMVFTSNFS--TIFSDIA 1507

Qy   350 SDKP----APMNLVKSEGVK-LPAPYQERTIDLSAYAGQQVILAFRHFNSTGIFRLYLDD 404
              ||| |||:||||:|:| |||||::||:||:||:||||::|||:||||||:|||:||
Db   1508 SVKPGGYQVQVNLADSSGNQAIPQTITINVVDTMGPVIKADNVSY-HVNTTKTEAEFFQD 1566

Qy   405 VAVSGEGGSNDYTYTVRDNVVIAQNLAATTNQENV---APGQYNYCVEVKYTAGVSPK 461
              ||:|| ||:|::|||||:|||:|:|||| ||:|:: ||||:||::|:|:|||:|
Db   1567 ARIDVTDNNDDTT------NLITSNFA------EKVNLNKPGKYEVTISATDTKG--NQ 1612

Qy   462 VCKDVTVEGSNE 473
              ||||:|||||:|
Db   1613 TTKEITVQVSKD 1624

RESULT 15
A36734
bacillopeptidase F (EC 3.4.21.-) precursor bpf [validated] - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 19-May-2000 #sequence revision 19-May-2000 #text change 16-Jun-2000
R;Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
J. Bacteriol. 172, 5520-5521, 1990
A;Reference number: A36734; A35131; A35750; B35750; S08223; JN0335; I39849; B69596; JU0084
R;Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
J. Bacteriol. 172, 5520-5521, 1990
A;Reference number: A35131; MUID:90170864; PMID:2106512
A;Accession: A35131
A;Molecule type: DNA
A;Residues: 1-365, 'S', 367-682, 'EIMP', 893, 'Q', 895-896 <SLO>
A;Cross-references: GB:M29035
A;Note: the authors translated the codon GAA for residue 545 as Leu
A;Accession: A36734
A;Molecule type: DNA
A;Residues: 1-1433 <SLO>
A;Cross-references: GB:M29035; NID:gi143307; PIDN:AAA62679.1; PID:gi143308
A;Contents: erratum
A;Accession: A36734

R;Wu, X.C.; Nathoo, S.; Pang, A.S.H.; Carne, T.; Wong, S.L.
J. Biol. Chem. 265, 6845-6850, 1990
A;Title: Cloning, genetic organization, and characterization of a structural gene encodi
A;Reference number: A35750; MUID:90216713; PMID:2108961
A;Accession: A35750
A;Molecule type: DNA
A;Residues: 1-392, 'V', 394-828, 'NIKTRLYSLKFCRSRHKSV' <WUA>
A;Cross-references: GB:J05400; NID:gi142607; PIDN:AAA83362.1; PID:gi142609
A;Note: this sequence has been corrected
A;Accession: B35750
A;Molecule type: DNA
A;Residues: 876-935,'CG' <WU2>
A;Cross-references: GB:J05400; NID:gi142607; PIDN:AAA83363.1; PID:gi1119197
A;Note: this sequence has been corrected
R;Masuda, E.S.; Anaguchi, H.; Sato, T.; Takeuchi, M.; Kobayashi, Y.
Nucleic Acids Res. 18, 657, 1990
A;Title: Nucleotide sequence of the sporulation gene spoIGA from Bacillus subtilis.
A;Reference number: S08223; MUID:90174995; PMID:2106671
A;Accession: S08223
A;Molecule type: DNA
A;Residues: 1410-1433 <MAS>
A;Cross-references: EMBL:X17344; NID:g40165; PIDN:CAA35224.1; PID:g809661
R;Kato, T.; Yamagata, Y.; Arai, T.; Ichishima, E.
Biosci. Biotechnol. Biochem. 56, 1166-1168, 1992
A;Title: Purification of a new extracellular 90-kDa serine proteinase with isoelectric p
A;Reference number: JN0335; MUID:93005071; PMID:1368833
A;Accession: JN0335
A;Molecule type: protein
A;Residues: 195-218,'A' <KAT>
A;Note: source of this material was Bacillus subtilis (natto)
R;Beall, B.; Lowe, M.; Lutkenhaus, J.
J. Bacteriol. 170, 4855-4864, 1988
A;Title: Cloning and characterization of Bacillus subtilis homologs of Escherichia coli
A;Reference number: I39846; MUID:89008108; PMID:3139638
A;Accession: I39849
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-211 <RES>
A;Cross-references: GB:M22630; NID:gi142938; PIDN:AAA22458.1; PID:gi551705
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte:
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch:
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle:
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel,
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Porcelli
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero:
akeuchi, M.; Takahashi, K.; Takemaru, K.; Takeuchi, M.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, I
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69596
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1433 <KUN>
A;Cross-references: GB:Z99111; GB:Z99112; NID:g2633902; GB:AL009126; NID:g2633902; PIDN:CAB13404.1; P:
A;Experimental source: strain 168
C;Genetics:
A;Gene: bpr; bpf
A;Map position: 135 (degrees)
C;Superfamily: bacillopeptidase F; subtilisin homology
C;Keywords: extracellular protein; hydrolase; serine proteinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-194/Domain: propeptide #status predicted <PRO>
F;195-1433/Product: bacillopeptidase F #status experimental <MAT>
F;218-466/Domain: subtilisin homology <SBT>
F;227,274,452/Active site: Asp, His, Ser #status predicted

Query Match   5.1%;   Score 135.5;   DB 1;   Length 1433;

```
Best Local Similarity   20.6%;   Pred. No. 1.1;
Matches  112;  Conservative   58;  Mismatches  150;  Indels  223;  Gaps   31;

Qy    90  ASIPEESF---WFANGTIPAGLYDPFEYKV------PVNADASFSPTNFVLDGTAS-----  136
              :::|:::   :|::||   ::|:|                       |
Db   437  SSVPGQTYEDGW--DGTSMAG---PHVSAVAALLKQANASLSVDEMEDILTSTAEPLTDS  491

Qy   137  --ADIPAGTYDYVIINFNPGIYIVGEGVSKGNDYVVEAGKT------YHFTVQRQGPGD  188
              ::  |  :  :  :::|                       |
Db   492  TFPDSPNNGYGHGLVNAFDA-VSAVTDGLGKAEGQVSVEGDDQEPPVVQHEKVTEAYEGG  550

Qy   189  AASVVVTGEGGNEFAPVQ------NLQWS------VSGQTVILTWQAPASD----KRTY-  231
              :  :|:|                :|:         ||:|:|::||     ||
Db   551  SLPITLTAEDNVSVTSVKLSYKLDQGEWTEITAKRISGDHLKGTYQAEIPDIKGTKLSYK  610

Qy   232  ----------VLNESFDT-------------------QTLPNGWTMIDADGDGHNWLSTINVNT  267
                           ::|                    ||:|||        :|:
Db   611  WMIHDFGHVVSSDVYDVTVKPSITAGYKQDFETAPGGWV---ASGTNNNW-------  658

Qy   268  ATHTGDGAMFSKSW-------TASGGAKI---DLSPDNY------LVTPKVTVPENGK  309
              ||: :   |        |:|::|       ::|:|          |:  ::|
Db   659  --------------EWGVPSTGPNTAASGEKVYGTNLT-GNYANSANMNLVMPPIKAPDSGS  705

Qy   310  LSYWVSSQVPWTNEH----YG-VFLSTTG----NEAANFTIKLLEETLGSDKPAPMNLVK  360
              |||        |         :||:              ::|   :|:
Db   706  LFLQFKS---WHNLEDDFDYGYVFVLPEGKNWEQAGVYNGK-----------------  744

Qy   361  SEGVKLPAPYQERTIDLSAYAGGQQVYLAFRHFNSTGIFR--LYLDDVAVSGEGSSNDYTY  418
              :  ||||||||       :  :||||:|:|||:      |||||:||||::||:|:  :
Db   745  -----TSSWTDEEIDLSAYKGGNIQVMFNLQSDESIAKEGWYIDDVVLSDKSAGK----  794

Qy   419  TVYRDNVVI----------------------AQNLAATTFNQENVA-----------  442
              ||: ::|                         ::|:  :|:|||
Db   795  TVKKNKLGVEKPSGKQKKPVNPKKAKPSANTAVKHQNKAIQPQVLPLKAQVSVVETGKS  854

Qy   443  ------PGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVG----QKVTL  491
                    ||| ::   :|:|:|                :|:           ||:||
Db   855  TYSDQSTGQYT----LKHKAG------DYT------------LMAEAYGYQSKTQKVSL  891

Qy   492  KWD  494
              ||
Db   892  KTD  894
```

Search completed: May 18, 2004, 11:47:45
Job time : 15.9765 secs

```
              GenCore version 5.1.6
     Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 18, 2004, 11:35:14 ; Search time 9.08355 Seconds
                                   (without alignments)
                                   2848.981 Million cell updates/sec

Title:        US-08-570-311-2
Perfect score: 2641
Sequence:     1 MRKLNSLFSLAVLLSLLCWG.........QNLTGSAVGQKVTLKWDAPN 497

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:     141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:  .   141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    SwissProt_42:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2641	100.0	2628	1	HGA2_PORGI	Q51845 porphyromon
2	2583	97.8	2164	1	HGA1_PORGI	P59915 porphyromon
3	494.5	18.7	991	1	CPG1_PORGI	P28784 porphyromon
4	472	17.9	989	1	PRTH_PORGI	P46071 porphyromon
5	156	5.9	1157	1	XYNA_THESA	P36917 thermoanaer
6	137.5	5.2	635	1	XYND_PAEPO	P45796 paenibacill
7	135.5	5.1	1433	1	SUBF_BACSU	P16397 bacillus su
8	135	5.1	1655	1	OMPB_RICCN	Q9kka3 r outer mem
9	131.5	5.0	725	1	YA33_SULSO	Q97297 sulfolobus
10	131	5.0	1656	1	OMPB_RICJA	O06653 r outer mem
11	129	4.9	794	1	OXAA_CHLCV	P59809 chlamydophi
12	127	4.8	2710	1	TOXA_CLODI	P16154 clostridium
13	124	4.7	1616	1	SLAP_BACCI	P35824 bacillus ci
14	121.5	4.6	972	1	CTA1_BACCI	P94286 bacillus ci
15	120.5	4.6	587	1	PICP_PSESR	P42790 pseudomonas
16	119	4.5	699	1	CHI1_BACCI	P20533 bacillus ci
17	118.5	4.5	1300	1	120K_RICRI	P14914 rickettsia
18	118.5	4.5	1654	1	OMPB_RICRI	Q53047 r outer mem
19	115.5	4.4	336	1	G3P1_SCHPO	P78958 schizosacch
20	114.5	4.3	335	1	G3PC_GRAVE	P54270 gracilaria
21	114.5	4.3	488	1	PHB_ALCFA	P12625 alcaligenes
22	114.5	4.3	681	1	TKT2_YEAST	P33315 saccharomyc
23	114.5	4.3	857	1	GELA_DICDI	P13466 dictyosteli
24	114.5	4.3	2660	1	YEEJ_ECO57	Q8x8v7 escherichia
25	113.5	4.3	1004	1	SLPO_BACBR	P09333 bacillus br
26	113.5	4.3	1151	1	ITA1_HUMAN	P56199 homo sapien
27	111	4.2	2201	1	TENA_HUMAN	P24821 homo sapien
28	111	4.2	2358	1	YEEJ_ECOLI	P76347 escherichia
29	110.5	4.2	408	1	G3PT_HUMAN	O14556 homo sapien
30	110.5	4.2	2265	1	FINC_BOVIN	P07589 bos taurus
31	110	4.2	642	1	FLID_CAMJE	Q9phw6 campylobact
32	110	4.2	827	1	CSG_HALVO	Q25062 halobacteri
33	110	4.2	1409	1	HAP1_HAEIN	P44596 haemophilus
34	109.5	4.1	932	1	CDG3_HUMAN	Q9y5h0 homo sapien
35	109.5	4.1	2021	1	OMPA_RICCN	Q52657 rickettsia
36	109	4.1	591	1	FLGE_CAUCR	P35806 caulobacter
37	109	4.1	917	1	SLAP_THETH	P35830 thermus the
38	109	4.1	1120	1	STFR_ECOLI	P76072 escherichia
39	108.5	4.1	337	1	G3P_PODAN	P32637 podospora a
40	108	4.1	611	1	WDR1_CAEEL	Q11176 caenorhabdi
41	108	4.1	828	1	MRKC_KLEPN	P21647 klebsiella
42	108	4.1	2481	1	FINC_XENLA	Q91740 xenopus lae
43	107.5	4.1	1475	1	APU_THETY	P16950 t amylopull
44	107.5	4.1	2477	1	FINC_RAT	P04937 rattus norv
45	107	4.1	666	1	NEP1_THEVU	Q60053 thermoactin

ALIGNMENTS

```
RESULT 1
HGA2_PORGI
ID   HGA2_PORGI     STANDARD;      PRT;   2628 AA.
AC   Q51845;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97047672; PubMed=8926061;
RA   Han N., Whitlock J., Progulske-Fox A.;
RT   "The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381
RT   contains four large, contiguous, direct repeats.";
RL   Infect. Immun. 64:4000-4007(1996).
CC   -!- FUNCTION: Agglutinates erythrocytes.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U41807; AAB17128.1; -.
DR   PIR; T28651; T28651.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25   2628       HEMAGGLUTININ A.
FT   DOMAIN      25    539       PEPTIDASE C25-LIKE 1.
FT   DOMAIN     540    995       PEPTIDASE C25-LIKE 2.
FT   DOMAIN     996   1451       PEPTIDASE C25-LIKE 3.
FT   DOMAIN    1452   1907       PEPTIDASE C25-LIKE 4.
FT   DOMAIN    2074   2628       PEPTIDASE C25-LIKE 5.
SQ   SEQUENCE   2628 AA;  283324 MW;  61C4DB32540C99DA CRC64;

Query Match         100.0%;  Score 2641;  DB 1;  Length 2628;
Best Local Similarity   100.0%;  Pred. No. 1.1e-162;
Matches  497;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy          1 MRKLNSLFSLAVLLSLLCWGTAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRKLNSLFSLAVLLSLLCWGTAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM 60

Qy         61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA 120
```

```
RESULT 2
HGA1_PORGI
ID   HGA1_PORGI     STANDARD;      PRT;   2164 AA.
AC   P59915;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA OR PG1837.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=22829867; PubMed=12949112;
RA   Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA   Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
RA   Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
RA   Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA   Dewhirst F.E., Fraser C.M.;
RT   "Complete genome sequence of the oral pathogenic bacterium
RT   Porphyromonas gingivalis strain W83.";
RL   J. Bacteriol. 185:5591-5601(2003).
CC   -!- FUNCTION: Agglutinates erythrocytes (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AE017178; AAQ66831.1; ALT_INIT.
DR   TIGR; PG1837; -.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
KW   Complete proteome.
FT   SIGNAL         1     25       POTENTIAL.
FT   CHAIN         26   2164       HEMAGGLUTININ A.
FT   DOMAIN        26    539       PEPTIDASE C25-LIKE 1.
FT   DOMAIN       540    991       PEPTIDASE C25-LIKE 2.
FT   DOMAIN       992   1443       PEPTIDASE C25-LIKE 3.
SQ   SEQUENCE   2164 AA;  233387 MW;  6DFAB2832586C63 CRC64;

  Query Match             97.8%;  Score 2583;  DB 1;  Length 2164;
  Best Local Similarity   98.0%;  Pred. No. 4.9e-159;
  Matches  487;  Conservative   0;  Mismatches   10;  Indels   0;  Gaps   0;

Qy     1  MRKLNSLFSLAVLLSLLCWGQTAAAQGPKTAPSVTHQAVQKGIRTSKVKVDLRDPIPAGM   60
          ||||||||||||||||||||||||||||||||||||||||||||||| |  |||||||||
Db     1  MRKLNSLFSLAVLLSLLCWGQTAAAQGPKTAPSVTHQAVQKGIRTSKAKDLRDPIPAGM   60

Qy    61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120

Qy   121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180

Qy   181  VQRQGPGDAASVVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  VQRQGPGDAASVVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240

Qy   241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300

Qy   301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAAANFTIKLLEETLGSDKPAPMNLVK  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAAANFTIKLLEETLGSDKPAPMNLVK  360

Qy   361  SEGVKLPAPYQERTIDLSAYAGGQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  SEGVKLPAPYQERTIDLSAYAGGQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420

Qy   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480

Qy   481  TGSAVGQKVTLKWDAPN  497
          |||||||||||||||||
Db   481  TGSAVGQKVTLKWDAPN  497


RESULT 3
CPG1_PORGI
ID   CPG1_PORGI     STANDARD;      PRT;    991 AA.
AC   P28784; Q45168;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
DE   (RGP-1).
GN   RGPA OR RGP1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 228-290 AND 517-541.
RC   STRAIN=381;
RX   MEDLINE=95168884; PubMed=7864651;
RA   Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
RA   Ikehara Y.;
RT   "Structural characterization of argingipain, a novel
RT   arginine-specific cysteine proteinase as a major periodontal
RT   pathogenic factor from Porphyromonas gingivalis.";
RL   Arch. Biochem. Biophys. 316:917-925(1995).
RN   [2]
RP   SEQUENCE OF 228-270.
RC   STRAIN=HG66;
RX   MEDLINE=92406812; PubMed=1527017;
RA   Chen Z., Potempa J., Polanowski A., Wikstrom M., Travis J.;
```

```
RT   "Purification and characterization of a 50-kDa cysteine proteinase
RT   (gingipain) from Porphyromonas gingivalis.";
RL   J. Biol. Chem. 267:18896-18901(1992).
CC   -!- FUNCTION: Thiol protease which is believed to participate in
CC       intracellular degradation and turnover of proteins. Its
CC       proteolytic activity is a major factor in both periodontal tissue
CC       destruction and in bacterial host defense mechanisms. Activates
CC       complement C3 and C5.
CC   -!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
CC       immunoglobulins, with a preference for Arg in P1, and hydrophobic
CC       residues in P2 and P3.
CC   -!- ENZYME REGULATION: Requires cysteine for activation and Ca(2+)
CC       and/or Mg(2+) for stabilization.It is stimulated by glycine-
CC       containing dipeptides. It is resistant to inhibition by proteinase
CC       inhibitors in human plasma.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   ---------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between  the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute.  There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26470; BAA05484.1; -.
DR   PIR; I40229; I40229.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
KW   Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
FT   SIGNAL        1     24
FT   PROPEP       25    227
FT   CHAIN       228    991       GINGIPAIN R1.
FT   ACT_SITE    438    438       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    471    471       NUCLEOPHILE (BY SIMILARITY).
FT   CONFLICT    264    265       RT -> TK (IN REF. 2).
SQ   SEQUENCE   991 AA;  108782 MW;  03BE3F43CEBBE2544 CRC64;

Query Match             18.7%;  Score 494.5;  DB 1;  Length 991;
Best Local Similarity   25.5%;  Pred. No. 2e-24;
Matches  124;  Conservative  41;  Mismatches  90;  Indels  231;  Gaps     7;

Qy       22 TAAAQG------------GPKTAPSVTHQAVQK-GIR---TSKVKDLRDPIPAGMARIILEAH   68
                                                       | | |||| ||||
Db      671 TATTGGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH  730

Qy       69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP  126
                |||  ||    |   |  | | :|| :|  | |   || ||
Db      731 DVWNDGSGYQILLDADHDQYGQVIPSDTHILW-FNCSVPANLFAPFEYTVPENADPSCSP  789

Qy      127 TNFVLDGTASADIPAGTGYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG  185
                |  | |    || || || |  |   | || |  || ||| |   | |  |   |||
Db      790 TNIMDGTASVNIPAGTGYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG  849

Qy      186 PGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQTLPNG  245
                  :  ::        | ::
Db      850 SGDGTELTIS------------------------------------------------  859

Qy      246 WTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVP  305
                                                                   | |||
Db      860 -----------------------------------------------------------  859

Qy      306 ENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVKSEGVK  365
               |  ||:         |     ||       | |  |  |
Db       58 GTEICCFPSLPKAPICSTSTLMRLRSKTNAKRADFTETFRSSTHGEAPAEWTTIDADGDG  117

Qy      366 LPAPYQERTIDLSAYAGAQVVLAFRHFNSTGIFRYLDDVAVSGEGSSNDYTYTVYRDNV  425
                                                            ||| |  |||
Db      118 QGWLCLSSGQLDWLTAHGGTNVVSSFSWNG-----MALNPDNYLISKDVTGATKVKYYA  172
```

```
Db      860 ------------------------------------------EGGGSDYTYTVYRDGT  875
                                                           ::   |: :|
Qy      426 VIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVBGSNEFAHVQNLTGSAV  485
                  ::   ||::  || :: : | ||||||  | |||:|:|||  |||   |::
Db      876 KIKEGLTETTYRDAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVV  935

Qy      486 GQKVTL  491
                ::||
Db      936 GKTITV  941
```

```
RESULT 4
PRTH_PORGI
ID   PRTH_PORGI     STANDARD;      PRT;    989 AA.
AC   P46071;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protease prtH (EC 3.4.22.-).
GN   PRTH.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroidetes (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=95012612; PubMed=7927685;
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RT   "Cloning and characterization of a new protease gene (prtH) from
RT   Porphyromonas gingivalis.";
RL   Infect. Immun. 62:4279-4286(1994).
RN   [2]
RP   ERRATUM.
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RL   Infect. Immun. 62:5707-5707(1994).
CC   -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
CC       P.GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
CC       IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
CC       AND IS A VIRULENCE FACTOR.
CC   -!- SUBCELLULAR LOCATION: In membrane vesicles.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   ---------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute.  There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L27483; AAA51298.1; -.
DR   HSSP; P23882; 1FMT.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR002376; formyl_transf.
DR   Pfam; PF00551; formyl_transf; 1.
KW   Hydrolase; Thiol protease; Repeat; Virulence.
FT   REPEAT      270    323
FT   REPEAT      528    581
SQ   SEQUENCE   989 AA;  110238 MW;  FA85FE8A3AC8944C CRC64;

Query Match             17.9%;  Score 472;  DB 1;  Length 989;
Best Local Similarity   38.6%;  Pred. No. 5.7e-23;
Matches  119;  Conservative  45;  Mismatches  124;  Indels  20;  Gaps     10;

Qy      199 GNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQT---LPNGWTMIDADGDG  255
                |||||| || || ||||| | |  |||| |||| |||||     ||| ||||||||
Db       58 GTEICCFPSLPKAPICSTSTLMRLRSKTNAKRADFTETFRSSTHGEAPAEWTTIDADGDG  117

Qy      256 HNW--LSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYW  313
                   ::|:|    |  |       :  :  ||:
Db      118 QGWLCLSSGQLDWLTAHGGTNVVSSFSWNG----MALNPDNYLISKDVTGATKVKYYYA  172
```

```
Qy   314  VSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDK-PAPMNL-VKSEGVKLPAPYQ 371
          |:   :|| ::|| |.||:|    :|: :|  .:||:.|:
Db   173  VNDGFP--GDHYAVMISKTGTNAGDFTVFEETPNGINKGGARPGLSTEADGAKPQSVWI 230

Qy   372  ERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTVYRDNVVIAQ 429
          |||:|| :|:|:|  |::  :||| |:||:|...  ::|::|.|||||:|:||||||:.:
Db   231  ERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTTTVYRDGTKIKE 287

Qy   430  NLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVITVEGSNEFAHVQNLTGSAVGQKV 489
          |||||||||||||||| |||.|:|||:|   ||  :|| |||||  |||:  :|:|:
Db   288  GLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFNPVKNLKAQPDGGDV 346

Qy   490  TLKWDAPN 497
          |||:|:|:
Db   347  VLKWEAPS 354
```

RESULT 5
XYNA_THESA
```
ID   XYNA_THESA     STANDARD;      PRT;   1157 AA.
AC   P36917;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) (Xylanase A)
DE   (1,4-Beta-D-xylan xylanohydrolase A).
GN   XYNA.
OS   Thermoanaerobacter saccharolyticum.
OC   Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC   Thermoanaerobacteriaceae; Thermoanaerobacterium.
OX   NCBI_TaxID=28896;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 34-41.
RC   STRAIN=B6A-RI;
RX   MEDLINE=94028940; PubMed=8215382;
RA   Lee Y.-E., Lowe S.E., Zeikus J.G.;
RT   "Gene cloning, sequencing, and biochemical characterization of
RT   endoxylanase from Thermoanaerobacterium saccharolyticum B6A-RI.";
RL   Appl. Environ. Microbiol. 59:3134-3137(1993).
RN   [2]
RP   ACTIVE SITE, AND MUTAGENESIS.
RX   MEDLINE=93388520; PubMed=8376336;
RA   Lee Y.-E., Lowe S.E., Henrissat B., Zeikus J.G.;
RT   "Characterization of the active site and thermostability regions of
RT   endoxylanase from Thermoanaerobacterium saccharolyticum B6A-RI.";
RL   J. Bacteriol. 175:5890-5898(1993).
CC   -!- FUNCTION: Endo-acting enzyme that randomly cleaves the internal
CC       xylosidic linkages of the xylan backbone, yielding
CC       xylooligosaccharides of various lengths which are further
CC       hydrolyzed to xylose molecules by beta-xylosidase (EC 3.2.1.37).
CC       The temperature and pH optima for this enzyme are 70 degrees
CC       Celsius and 5.5, respectively. Requires at least three xylose
CC       residues for catalytic activity. Does not have activity against
CC       xylobiose.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
CC       linkages in xylans.
CC   -!- PATHWAY: Xylan degradation.
CC   -!- INDUCTION: By xylan and xylose.
CC   -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC       hydrolases).
CC   -!- SIMILARITY: Contains 2 S-layer homology (SLH) domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M97882; AAA21812.1; ALT_SEQ.
DR   HSSP; P10478; 1XYZ.
DR   InterPro; IPR008960; CBD9-like.
DR   InterPro; IPR003305; CBM_CenC.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR001000; Glyco_hydro_10.
DR   InterPro; IPR001547; Glyco_hydro_5.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF02018; CBM_4_9; 2.
DR   Pfam; PF00331; Glyco_hydro_10; 1.
DR   Pfam; PF00395; SLH; 2.
DR   PRINTS; PR00134; GLHYDRLASE10.
DR   SMART; SM00663; Glyco_10; 1.
DR   PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR   PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
KW   Xylan degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT   SIGNAL         1       33
FT   CHAIN         34     1157       ENDO-1,4-BETA-XYLANASE A.
FT   DOMAIN       413      649       CATALYTIC (POTENTIAL).
FT   DOMAIN      1055     1113       SLH 1.
FT   DOMAIN      1114     1157       SLH 1.
FT   ACT_SITE     495      495       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE     537      537       NUCLEOPHILE.
FT   ACT_SITE     600      600       D-N: LOSS OF ACTIVITY.
FT   MUTAGEN      537      537       D->N: LOSS OF ACTIVITY.
FT   MUTAGEN      600      600       E->Q: LOSS OF ACTIVITY.
FT   MUTAGEN      602      602       D->N: LOSS OF ACTIVITY.
SQ   SEQUENCE   1157 AA;   128379 MW;   51FA600497EC58B CRC64;
```

```
Query Match             5.9%;  Score 156;  DB 1;  Length 1157;
Best Local Similarity   20.4%;  Pred. No. 0.019;
Matches   92;  Conservative   64;  Mismatches  128;  Indels  166;  Gaps   23;
```

```
Qy   106  AGLYDPFEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSK 165
               :|  |:   |:| |  : :|:   :
Db    22  ASLFSP-----PIRVFADDTNINILVSNG---DFESGTIDGWIKQGNPTLAVTTEQAIGQ 72

Qy   166  GNDYVVEAGKTYHFTVQRQGP----------VVTGE-GGNEPAPVQNLQ 209
               |::|| :  :|::: .|||          |:|::  ::|.|.|:
Db    73  YSMKVTGRTQTY------EGPAYSFLGKMQKGESYSVSLKVRLVSGQNSSNPLITVTMFR 126

Qy   210  WSVSGQTV-TLTWQAPASD--------------------KRTYVLNESFD------ 238
               :|| :|  |:.|:|.:|                    |:|| :::|
Db   127  EDDNGKHYDTIVWQKQVSEDSWTTVSGYTLDYIGTLKTLYMYVESPDPTLEYYIDDVVV 186

Qy   239  --------------TQTLPNGWTMIDADGGHNWLSTIN----VYNTATHTGDGAMFSKS 280
                             | ||:|  |: .:| |:  |:     |:  :| .|:.:|:
Db   187  TTQNPIQVGNVIANETFPNGNT-----------SGWIGTGSSVVKAVYGVA-HSGDYSLLTTG 237

Qy   281  WTAS-GGAKIDLSPDNYLVTPKVTVPENGKLSTWVSSQVPWTNEHYGVFLSTTGNEAANF 339
               ||:  |.||.| ||||:|||||||:|                    ::
Db   238  RTANWNGPSYDL-----------TGKIVPGQQVNVDFWVK----------FVNGNDTEQIKA 278

Qy   340  TIKLLEETLGSDKPAPM------NLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFN 393
               |:|  :| :|:|  |:      :|:: :  |: :|  ||  |:|.: |.::|:
Db   279  TVKAT-----SDKDNYIQVNDFANVNKGEWTELKGSF---TLPVADYSGISIVESQ-N 328

Qy   394  STGIFRLYLDDVAVSGEGSSNDYT---------YTVYRD------------------- 423
               :|.|  : |::| :|..::|.:|         ||:|:
Db   329  PT--LEFYIDDFSVIGEISNNQITIQNDIPDLYSVFKDYFPIGAVDPSRLNDADPHAQL 386

Qy   424  ----NVVIAQNLAATTFNQENVAPGQYNY 448
               :|:  |:|| :|::|::|.|||:|
Db   387  TAKFNFMLVAEN---AMKPESLQPTEGNF 412
```

RESULT 6
XYND_PAEPO
```
ID   XYND_PAEPO     STANDARD;      PRT;   635 AA.
AC   P45796;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Endo-1,4-beta-xylanase D precursor (EC 3.2.1.8) (Xylanase D)
DE   (1,4-beta-D-xylan xylanohydrolase D).
```

GN XYND.
OS Paenibacillus polymyxa (Bacillus polymyxa).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1406;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 842;
RX MEDLINE=92041687; PubMed=1938968;
RA Gosalbes M.J., Perez-Gonzalez J.A., Gonzalez R., Navarro A.;
RT "Two beta-glycanase genes are clustered in Bacillus polymyxa:
RT molecular cloning, expression, and sequence analysis of genes
RT encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase.";
RL J. Bacteriol. 173:7705-7710(1991).
CC -!- FUNCTION: SHOWS XYLANASE ACTIVITY AS WELL AS ALPHA-L-
CC   ARABINOFURANOSIDASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
CC   linkages in xylans.
CC -!- PATHWAY: Xylan degradation.
CC -!- SIMILARITY: Belongs to family 43 of glycosyl hydrolases.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; X57094; CAA40378.1; -.
DR PIR; S19011; S19011.
DR InterPro; IPR006584; CBD_IV.
DR InterPro; IPR005084; CBM_6.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF03422; CBM_6; 1.
DR SMART; SM00606; CBD_IV; 1.
KW Xylan degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 635 ENDO-1,4-BETA-XYLANASE D.
SQ SEQUENCE 635 AA; 67914 MW; F9DEC6996732331G CRC64.

Query Match 5.2%; Score 137.5; DB 1; Length 635;
Best Local Similarity 21.2%; Pred. No. 0.14;
Matches 123; Conservative 65; Mismatches 176; Indels 217; Gaps 32;

```
Qy     22 TAAAQGGPKTAP---SVTHQAVQKGIRTSKVKDLRDPIPAGMARIILEAHDVWEDGTGY  77
              ||||           ||||
Db    156 TADTPIGPWTDPLGKALVTHST-----------------PGNAGVTWLFDPAVLVDDDGTGY 200

Qy     78 QMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKV------------PVNADA 122
             ::|                                   ::|
Db    201 --LYS------GGGIPNES------DPASIANPKTARVIKLGADMTSVIGSATTIDAPY 245

Qy    123 SFSPT-----------------NFVLDGTASADIPAGTYDYVII-NP---------NPG 154
             ||||                 ||              |::
Db    246 LFEDSGIHKYNGKYYSYCINFA--GTHPQQYPAGEIGYMVSDNPMGPFTYKGHFLKNPY 303

Qy    155 IIYIVG-----------------EGVSKGNDYVVEAGKTY---HFT-VQRQGPGDAA 190
             ||                    | ||||   :|||||
Db    304 TFFGVGGNNHHAVFNFKNEWYVVVHAQTVSKAQ---ICAGKGYRSPHINKLVHKEDGSIS 360

Qy    191 SVV--VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQTLPNGWTM 248
             :|                      ||||
Db    361 EVQGNMTG-----IAQLSNMNPYTRVEAETIAWQAGVTTEPT----------------- 397

Qy    249 IDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAK--------------IDLSP 293
             ||||                ||||        |                       :|
Db    398 -QASGGP---ISNLNVTN--IHNGDWIAVGKADFGSAGAKTFKANVATNVGGNIEVRLDS 451

Qy    294 DNYLVTPKVTVPENGKLSYW--VSSQVPWTNEHYGVFLSTTGNEAAN-----------FTIKL 343
             ::||                 |::                         :|:||||
Db    452 ETGPLVGSLKVPSTGGMQTWREVTTINNATGVHNIYIVFTGSGSGNLLNLDAWQFTPNT 511

Qy    344 LEETLGSDKPAPMNLVKSBGVKLPAPYQERTIDLSAYAGQQVVLAFRHF--NSTGIFRLY 401
```

Db 512 GGNTITKVEAENMKIGGTYAGKISAPFD---GVALYANAD-YVSYSQYFANST------ 560
    |:| :::|| || |:::|| ||
Qy 402 LDDVAVSGEGSSNDYTYTYRDNVVIAQNLAATFNQENVAPGQYNYCVEVKYTAGVSPK 461
    :::|| :|| |||| :::
Db 561 -HNISVRG-ASSNAGTAKV---DIVIG------GVTVGSFNF------TGKTPT 597
       :| :|:||
Qy 462 VCKDVTVBGSNEFAHVQNLTGSAVGQKVTL------KWDA 495
      ||| |:|: |||
Db 598 V------------QTLSNTHATGDQEIKLALTSDDGTWDA 626

RESULT 7
SUBF_BACSU
ID SUBF_BACSU   STANDARD;   PRT; 1433 AA.
AC P16397;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bacillopeptidase F precursor (EC 3.4.21.-) (Esterase) (RP-I protease)
DE (90 kDa serine proteinase).
GN BPR OR BPF OR BSU15300.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 195-222.
RX MEDLINE=90170864; PubMed=2106512;
RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
RA Pero J.;
RT "Bacillopeptidase F of Bacillus subtilis: purification of the protein
RT and cloning of the gene.";
RL J. Bacteriol. 172:1470-1477(1990).
RN [2]
RP REVISIONS.
RC STRAIN=168;
RX MEDLINE=90368623; PubMed=2118514;
RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
RA Pero J.;
RL J. Bacteriol. 172:5520-5521(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90216711; PubMed=2108961;
RA Wu X.-C., Nathoo S., Pang A.S.-H., Carne T., Wang S.-L.;
RT "Cloning, genetic organization, and characterization of a structural
RT gene encoding bacillopeptidase F from Bacillus subtilis.";
RL J. Biol. Chem. 265:6845-6850(1990).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,

```
FT   CONFLICT       844    852        AQVSVVETG -> FCRSRHKSV (IN REF. 3).
FT   CONFLICT       853   1433        MISSING (IN REF. 3).
SQ   SEQUENCE      1433 AA;  154577 MW;  9BDF6846897807C9 CRC64;

Query Match                    5.1%;  Score 135.5;  DB 1;  Length 1433;
Best Local Similarity         20.6%;  Pred. No. 0.54;
Matches  112;  Conservative  58;  Mismatches 150;  Indels 223;  Gaps  31;

Qy     90 ASIPEESF---WFANGTIPAGLYDPFEYKV----PVNADASFSPTNFVLDGTAS----- 136
          :|:  ||| :        ||         ||       |
Db    437 SSVPGQTYEDGW--DGTSMAG--PHVSAVAALLKQANASLSVDEMEDILTSTAEPLTDS 491

Qy    137 --ADIPAGTYDVIINPNPGIIYIVGEGVSKGNDYVVEAGKT------YHFTVQRQGPGD 188
             :                 |         ::::
Db    492 TFPDSPNNGYGHGLVNAFDA-VSAVTDGLGKAEGQVSVEGDDQEPPVYQHEKVTEAYEGG 550

Qy    189 AASVVTGEGGNEFAPVQ------NLQWS------VSGQTVLTWQAPASD----KRTY--- 231
          ::  :|                      :|                 :|:
Db    551 SLPLTLTAEDNVSVTSVKLSYKLDQQEWTEITAKRISGDHIKGTYQAEIPDIKGTKLSYK 610

Qy    232 ------------VLNESFDT---------------------QTLPNGWTMIDADGDGHNWLSTINVYNT 267
                      ::.:                        ::  :|
Db    611 WMIHDFGGHVVSSDVYDVTVKPSITAGYKQDFETAPGGWV---ASGTNNNW-------- 658

Qy    268 ATHTGDGAMFSKSW-----------TASGGAKI--DLSPDNY------LVTPKVTVPENGK 309
                      |       ||:|::|         |:.|      || :|.
Db    659 -----------------EWGVPSTGPNTAASGEKVYGTNLT-GNYANSANMNLVMPPIKAPDSGS 705

Qy    310 LSYWVSSQVPWTNEH---YG-VFLSTTG---NEAANFTIKLLEETLGSDKPAPMNLVK 360
          |            |      |    |      :|      :| |
Db    706 LFLQFKS---WHNLEDDFDYGYVFVLPEGEKNWEQAGVYNGK------------------ 744

Qy    361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFR--LYLDDVAVSGEGSSNDYTY 418
                  |||     |||:||    :|   |  ||:|:|
Db    745 ------TSSWTDEEIDLSAYKGQNIQVMFNLQSDESIAKEGWYIDDVVLSDKSAGK---- 794

Qy    419 TVYRDNVVI-------------------AQNLAATTFNQENVA----------------- 442
          ||:|::|                    |   |:|
Db    795 TVKKNKLGVEKPSGKQKKPVNPKKAKPSANTAVKHQNKAIQPQVLPLKAQVSVVETGKS 854

Qy    443 ------PGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVG----QKVTL 491
                 ||| || :||   |||      |  |:|      | |  |  |       ::.:
Db    855 TYSDQSTGQYT---LKHKAG------DYT----------LMAEAYGYQSKTQKVSL 891

Qy    492 KWD 494
          |
Db    892 KTD 894
```

```
RESULT 8
OMPB_RICCN
ID   OMPB_RICCN    STANDARD;       PRT;  1655 AA.
AC   Q9KKA3; Q9KK98; Q9XC45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB OR RC1085.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Malish 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
```

```
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [5]
RP   SEQUENCE OF 1-211 FROM N.A.
RX   MEDLINE=89008108; PubMed=3139638;
RA   Beall B., Lowe M., Lutkenhaus J.;
RT   "Cloning and characterization of Bacillus subtilis homologs of
RT   Escherichia coli cell division genes ftsZ and ftsA.";
RL   J. Bacteriol. 170:4855-4864(1988).
RN   [6]
RP   SEQUENCE OF 1410-1433 FROM N.A.
RC   STRAIN=168 / Marburg;
RX   MEDLINE=90174995; PubMed=2106671;
RA   Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;
RT   "Nucleotide sequence of the sporulation gene spoIIGA from Bacillus
RT   subtilis.";
RL   Nucleic Acids Res. 18:657-657(1990).
RN   [7]
RP   SEQUENCE OF 195-219.
RC   STRAIN=Natto 16;
RA   Kato T., Yamagata Y., Arai T., Ichishima E.;
RT   "Purification of a new extracellular 90-kDa serine proteinase with
RT   isoelectric point of 3.9 from Bacillus subtilis (natto) and
RT   elucidation of its distinct mode of action.";
RL   Biosci. Biotechnol. Biochem. 56:1166-1168(1992).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to peptidase family S8.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M29035; AAA62679.1; -.
DR   EMBL; J05400; AAA83362.1; -.
DR   EMBL; Z99111; CAB13403.1; -.
DR   EMBL; M22630; AAA22458.1; -.
DR   EMBL; X17344; CAA35224.1; -.
DR   PIR; A36734; A36734.
DR   HSSP; P00782; 2SBT.
DR   MEROPS; S08.017; -.
DR   SubtiList; BG10233; bpr.
DR   InterPro; IPR008757; Peptidase_M6.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR009020; Protease_inhib.
DR   Pfam; PF05547; Peptidase_M6; 1.
DR   Pfam; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
FT   SIGNAL         1     30
FT   PROPEP        31    194        POTENTIAL.
FT   CHAIN        195    755        BACILLOPEPTIDASE F.
FT   PROPEP       756   1433        POTENTIAL.
FT   ACT_SITE     227    227        CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE     274    274        CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE     452    452        CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CONFLICT     219    219        T -> A (IN REF. 7).
FT   CONFLICT     393    393        A -> V (IN REF. 3).
FT   CONFLICT     829    834        KHQNKA -> N (IN REF. 3).
FT   CONFLICT     836    841        QPQVLP -> RTRLYS (IN REF. 3).
```

```
RL   Science 293:2093-2098(2001).
RN   [2]
RP   SEQUENCE OF 33-1649 FROM N.A.
RC   STRAIN=Indian tick typhus, and Malish 7;
RX   MEDLINE=20393643; PubMed=10933649;
RA   Roux V., Raoult D.;
RT   "Phylogenetic analysis of members of the genus Rickettsia using the
RT   gene coding the outer-membrane protein rOmpB (ompB).";
RL   Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN   [3]
RP   SEQUENCE OF 353-1655 FROM N.A.
RC   STRAIN=Malish 7;
RA   Stenos J., Walker D.;
RT   "The rickettsial outer membrane protein A and B genes of Rickettsia
RT   australis, the most divergent rickettsia of the spotted fever group.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC       similarity).
CC   -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC       (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC       layer with hexagonal symmetry (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AE008659; AAL03623.1; -.
DR   EMBL; AF123721; AAF34124.1; -.
DR   EMBL; AF123726; AAF34129.1; -.
DR   EMBL; AF149110; AAD39533.1; -.
DR   PIR; E97835; E97835.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   Pfam; PF03797; Autotransporter; 1.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 2.
KW   Antigen; S-layer; Cell wall; Complete proteome.
FT   CHAIN        1   1334       120 kDa SURFACE-EXPOSED PROTEIN.
FT   CHAIN     1335   1655       32 kDa BETA PEPTIDE.
FT   VARIANT     61     61       P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     75     75       G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     78     78       K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT    251    251       V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT    413    413       N -> V (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT    959    959       I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT    988    988       A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT   1139   1139       R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT   CONFLICT   353    354       KD -> GH (IN REF. 3).
FT   CONFLICT   776    776       F -> S (IN REF. 3).
FT   CONFLICT  1159   1159       E -> D (IN REF. 3).
FT   CONFLICT  1177   1177       G -> R (IN REF. 3).
FT   CONFLICT  1492   1492       H -> G (IN REF. 3).
SQ   SEQUENCE  1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;

Query Match             5.1%;  Score 135; DB 1; Length 1655;
Best Local Similarity  19.3%;  Pred. No. 0.7;
Matches  102;  Conservative  53;  Mismatches  189;  Indels  184;  Gaps  20;

Qy  100 ANGTIPAGLYDPFEYKVPVNADASFSPTNFV------------------LDGTASADIPAGTYDYV 147
           |||| ||                              :|| ||
Db  439 ANGTLAS------------ASADANVAVTNNITAIEASGAGVVQLSGTHRAEBLRLG------ 482

Qy  148 IINPNPGIYIYIVGEGV--SKGNDYVVEAGKTYHFTVQRQGPGDAASVVTGEGNE--F 202
           :||   ||              |:   ||
Db  483 ---NAGSVFKLADGTVINGKVNQTALVGGALAAGTITLDG----SATITGDIGNAGGA 533
```

```
Qy  203 APVQ------------------------------NLQWSVSGQTVLTWQ------------APASD 227
           :|
Db  534 AALQGITLANDATKTLTLGGANIIGANGGTINFQANGGTIKLTSTQNNIVVDFDLAIATD 593

Qy  228 KRTYVLNESFDTQTLPNGWTM-----IDADGDGHNWLSTINVYNTATHTGDGAMFSKSWT 282
           ||                                :||:  :|:
Db  594 QTGVV-----DASSLTNAQTLTINGKITGTVGANNKTLGQFNIGSSKKTVLSDGVAINELV 648

Qy  283 ASGGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIK 342
Db  649 IGNNGAVQFAHNTYLITRTTNAAGQGKIIF-----NPVVNNNTTL--ATGTNLGSATNP 700

Qy  343 LLEETLGSDKPAPMNLVKS--EGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRL 400
           ||   :   ||           ||:   :|
Db  701 LARINFGSKGAANVDTVLNVGKGVNL---YATNITTTDANVGSFIFNA----GGTNI--- 750

Qy  401 YLDDVAVSGE--GSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTA- 456
           :|||                |:    :|
Db  751 -----VSGTVGQQGNKFNTVALDNGTTVKFLGNATFNGNTTIAANSTLQIGGNYTADF 804

Qy  457 --------------GVSPKVCKDVTVEGS-----NEFAHVQNLTG----- 482
                        ||:|
Db  805 VASADGTGIVEFVNTGPITVTLNKQAAPVNALKQITVSGPGNVVINEIGNAGNYHGAVTD 864

Qy  483 --------SAVGQKVTLKWDAPN 497
                   |||
Db  865 TIAFENSSLGAVVFLPRGIPFNDAGNRIPLTIKSTVGNKTATGFDVPS 912


RESULT 9

ID   YA33_SULSO     STANDARD;      PRT;   725 AA.
AC   Q97297;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein SSO1033 precursor.
GN   SSO1033.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=1427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC   -!- SIMILARITY: Contains 6 fibronectin type III domains.
CC   -!- SIMILARITY: Contains 6 Kelch repeats.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AE006722; AAK41296.1; -.
DR   PIR; A90255; A90255.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR006652; Kelch_rep.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF01344; Kelch; 5.
```

```
DR   SMART; SM00060; FN3; 3.
KW   Hypothetical protein; Kelch repeat; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    725       HYPOTHETICAL PROTEIN SSO1033.
FT   REPEAT       59    100       KELCH 1.
FT   REPEAT      101    145       KELCH 2.
FT   REPEAT      146    199       KELCH 3.
FT   REPEAT      201    248       KELCH 4.
FT   REPEAT      250    297       KELCH 5.
FT   REPEAT      299    342       KELCH 6.
FT   DOMAIN      323    398       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      412    490       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      585    652       FIBRONECTIN TYPE-III 3.
SQ   SEQUENCE   725 AA;  78465 MW;  FE8F1220D013BAB6 CRC64;

Query Match               5.0%;   Score 131.5;  DB 1;  Length 725;
Best Local Similarity    20.4%;   Pred. No. 0.4;
Matches   112;  Conservative  61;  Mismatches  134;  Indels  241;  Gaps  27;

Qy    101  NGTIPAGLYDPFEYKVPVNADASFSPTN-----------FVLDGTASADI----  139
Db    154  NTTSPAGLYFP-----PSNAIRLFYPNNDSWRIIGYMPVPTYGGGYVFNGTSLLIIVSGYI  208

Qy    140  --PAGTYDVIINPN-------------------------PGIIYIVGEGVSKGNDYVVEAG  174
Db    209  GYSAYTNDILIYSPQNNNWTILNGVLPYWIHDSALAYYRGVLFIVG-----GYIYTAG     261

Qy    175  K-------TYHFTVQRQG------PGDAASVV------VTGEGGNEFAPVQNLQ-------  209
Db    262  SGGVNNAILAYYNGNLQRVGYLPVPVYSAGYVQVGNMLYLAGGIGSSLSDVSALQLITFN   321

Qy    210  -----------WSVSGQTVLTWQAPASDKRTYVLNESFDTQTLPNGWTMIDADGDGHNWL--  259
Db    322  FPPLPPKITSYSAGNESVTLGW--------------------NPVRLSSGYEIIYWNNMGFN--  363

Qy    260  STINVYNTATHT---GDGAMF-------------------------------             277
Db    364  SSINVGNVTSYTVTGLKDGITYYFEVLAYNSIGYSSPSSIIALTPASVPNPPQLVSVKYG   423

Qy    278  ---SKSW----TASG------GAKIDLSPDNYLV---------------------TPKVT-----  303
Db    424  NDNVTLNWLPPTPSGGYLLLGYYVIVKNENSMVSSHFVNSTSLTISNLTPNVTYNVFIYA    483

Qy    304  VPENGKLSYWVSSQVPWTNEHYGVFLSTGNE-AANFTIKLLEETLGSDKPAPMNLVKSE    362
Db    484  VNKLGNSSPLVLTVVPITKASVFAFITKLGNGILVNWTTSF-------PANIT--       529

Qy    363  GVKLPAPYQERTIDLSAYAGQQYLAFR----------HFNSTGIFR------LYL--       402
Db    530  -LELYNPNGNLISQIAIKGNSSYL-FRVPQGNYTLVIIASNSAGVSKYVYQVVYYLPPA   587

Qy    403  ----------DDVAVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEV    452
Db    588  SPQVSLIGFGNNLYISWNNEAANVITYLVYVNNSLVYEGPSNSIVT--NISNG--TYLVKV   643

Qy    453  KYTAGVSP  460
Db    644  --IGVNP.  648

RESULT 10
OMPB_RICJA
ID   OMPB_RICJA   STANDARD;   PRT;   1656 AA.
AC   O06653;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rompB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB.
OS   Rickettsia japonica.
```

```
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=35790;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YH;
RA   Uchiyama T.;
RT   "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT   japonica.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC       similarity).
CC   -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC       (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC       layer with hexagonal symmetry.
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB003681; BAA20138.1; -.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   Pfam; PF03797; Autotransporter; 1.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 2.
KW   Antigen; s-layer; Cell wall.
FT   CHAIN         1   1338       120 kDa SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1339   1656       32 kDa BETA PEPTIDE.
FT   DOMAIN      528    533       POLY-GLY.
SQ   SEQUENCE   1656 AA;  168097 MW;  3132A69C9DD5999F CRC64;

Query Match               5.0%;   Score 131;  DB 1;  Length 1656;
Best Local Similarity    21.3%;   Pred. No. 1.3;
Matches    99;  Conservative  45;  Mismatches  198;  Indels  122;  Gaps  18;

Qy    100  ANGTIPAGLYDPFEYKVPVNADASFSPTNFV-----------LDGTASADIPAGT----  143
Db    439  ANGTILAS------------ASADANVAVTNNITAIEASGVGVQLSGTHTAELRLRIGNAGSV  487

Qy    144  ---YDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFTVQRQGPG-----------DAA  190
Db    488  FKLADGTVINGKVNQTVLVG-GVLAAGAITLDGSATTTGDIGNGGGGAALQSITLANDAT  546

Qy    191  SVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQ-----------APASDKRTYVLNESFDTQ  240
Db    547  KTLTLGGANIISANGGTINFQANGGTIKLTSTQNNIVDCDLAIATDQTGVV------DAS  601

Qy    241  TLPNGWTM-----IDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN  295
Db    602  SLTNAQTLTISGTIGIIGANNTTLGQNFIGSSKTTLNGGNVAINELVIGNNGSVQFAHNT  661

Qy    296  YLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDKRPAP  355
Db    662  YLITRTTNAAGQQKIIF---NPVVNNTTLAAGTNLGSAAN--PLAEINFGS-KGAR     712

Qy    356  MNLVKS--EGVKLPAPYQERTIDLSAYAGQQYLAFRHFNSTGIFRLYLDDVAVSGE--G  411
Db    713  ADTVLNVGEGVNL--YATNITTTDANVGSFV---FNAGG-------KNIVSGTVGG     756

Qy    412  SSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTA-------------  456
Db    757  QQGNKFNTVALDNGTTVKFLGNATFNGNTTIAANSTLQISGNTADFIASADGTGVEFV  816

Qy    457  -------GVSPKVCKDVTVEGS------NEFAHVQNLTGS  483
```

```
Db      817 NTGPINVTLNKQAVPVNALKQITVSGPGNVVNEIGNAGNYHGA 860

RESULT 11
OXAA_CHLCV
ID   OXAA_CHLCV    STANDARD;    PRT;   794 AA.
AC   P59809;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Inner membrane protein oxaA.
GN   OXAA OR CCA00472.
OS   Chlamydophila caviae.
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83557;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GPIC;
RX   MEDLINE=22569155; PubMed=12682364;
RA   Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA   Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,
RA   Umayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA   Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,
RA   Fraser C.M.;
RT   "Genome sequence of Chlamydophila caviae (Chlamydia psittaci GPIC):
RT   examining the role of niche-specific genes in the evolution of the
RT   Chlamydiaceae.";
RL   Nucleic Acids Res. 31:2134-2147(2003).
-!- FUNCTION: Required for the insertion of integral membrane proteins
    into the membrane. Probably plays an essential role in the
    integration of proteins of the respiratory chain complexes.
    Involved in integration of membrane proteins that insert
    dependently and independently of the Sec translocase complex (By
    similarity).
-!- SUBUNIT: Specifically interacts with transmembrane segments of
    nascent integral membrane proteins during membrane integration (By
    similarity).
-!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
    (By similarity).
-!- SIMILARITY: Belongs to the OXA1/oxaA family. Subfamily 1.
--------------------------------------------------------------------------
```
```
DR   EMBL; AE016995; AAP05217.1; -.
DR   TIGR; CCA00472; -.
DR   HAMAP; MF_01810; -; 1.
DR   InterPro; IPR001708; 60kDa_innermemb.
DR   Pfam; PF02096; 60KD_IMP; 1.
KW   Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM      5     24       POTENTIAL.
FT   TRANSMEM    571    593       POTENTIAL.
FT   TRANSMEM    639    661       POTENTIAL.
FT   TRANSMEM    691    713       POTENTIAL.
FT   TRANSMEM    740    762       POTENTIAL.
SQ   SEQUENCE   794 AA;  88951 MW;  9EDCF05E73709133 CRC64;

Query Match          4.9%;  Score 129; DB 1; Length 794;
Best Local Similarity   21.0%;  Pred. No. 0.65;
Matches   86;  Conservative   63;  Mismatches  154;  Indels 106;  Gaps  21;

Qy     80 LWDADHNQYGASIPEESFW-FANGTIPAGLYDPFEYK-----VPVNADASFSPTNFVLDG 133
             ::       |        |:|| :||| |::| |:     ||: : :     ||| :
Db     80 VYNKDSAVYGTSL---VFWRSGNEYLPLGIYNSKEERLESLDLPITKAAVFSDSKSANAG 223

Qy    134 TASAD--IPAGTYDVIINPNPGIIYIVEGV-----SKGNDYVV-EAGKTYHFTVQRQG 185
            :||     |||  |:  ||    :::| |:        ||  |      ||   | |:
Db    167 VYNKDSAVYGTSL---VFWRSGNEYLPLGIYNSKEERLESLDLPITKAAVFSDSKSANAG 223

Qy    134 TASAD--IPAGTYDVIINPNPGIIYIVEGV-----SKGNDYVV-EAGKTYHFTVQRQG 185
Db    224 ANSAQYFVLSNEYMQLIVSQESGSI----EGINLPFSSEDNKSIVNEIG--FDRELKAQV 277
```

```
Db    186 PGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASD--KRTYVLNESFDTQTLP 243
              :::   :|  ||  :  ||  :|||:||  |::|| |:||     ||:|||||    :
Qy    278 PSSASFPGLPSVGANN-QPVSD---TVGGYYPLLRRGILSDAKKRT--PSSYHALNIV 329

Db    244 NGWTMIDADGDGHNWLSTINVYNTATHTGDGAM--------------------FSKS 280
              ::  :|:     |::  || ||| : :|:  |:|::
Qy    330 SGRELTNSVASGYR-VSTFNSTMLELESNDGSIKKTKTYKLPQQQPYAFFEVEGVNRASDDL 388

Db    281 WTASGGAKIDLSPDNYLVTP-----------------------KVTVPE----------- 317
              ||:| | |  |  |: ||                        ||:||
Qy    389 WITSGIPEVEIMSNAF--TPAIKYHVIKKNKGGLDKVKLPKAKDPLALRSGVVPQWILN- 445

Db    318 VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVKSEG---------VK 365
              :||::||||  | ||:   :  ::|   |     ||:   |  ||           |:
Qy    446 ---SNGYFGIILSPLTDIPAGYAASYVP---GSSVPTRLSLLSPKNQAYPASKYPGYETL 499

Db    366 LPAPYQERTIDLSAYAGQOVYLAFR------HPNSTGIFRLYLDDVAVSG 409
              |     ::|| ::|  :|  |||:|         ||  : |:||||:||
Qy    500 LPLPQKEGTHRFLVYAGPLADPTLRVLDKAYTNSKGESPQYLDCITFRG 548
```

```
RESULT 12
TOXA_CLODI
ID   TOXA_CLODI    STANDARD;    PRT;  2710 AA.
AC   P16154;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Toxin A.
GN   TOXA OR TCDA.
OS   Clostridium difficile.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1496;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RX   MEDLINE=90221894; PubMed=2109310;
RA   Sauerborn M., von Eichel-Streiber C.;
RT   "Nucleotide sequence of Clostridium difficile toxin A.";
RL   Nucleic Acids Res. 18:1629-1630(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RX   MEDLINE=90129305; PubMed=2105276;
RA   Dove C.H., Wang S.Z., Price S.B., Phelphs C.J., Lyerly D.M.,
RA   Wilkins T.W., Johnson J.L.;
RT   "Molecular characterization of the Clostridium difficile toxin A
RT   gene.";
RL   Infect. Immun. 58:480-488(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RA   von Eichel-Streiber C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
-!- FUNCTION: Only after the enteral delivery of the enterotoxin A may
    the characteristic disease called pseudomembranous colitis be
    induced.
-!- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
    REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
    DIFFERENT OLIGOPEPTIDES.
--------------------------------------------------------------------------
```
```
DR   EMBL; X51797; CAA36094.1; -.
DR   EMBL; M30307; AAA23383.1; -.
```

RL   Gene 147:1-11(1994).
CC   -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC   OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. MAY PLAY A ROLE
CC   IN THE EXPORT OF BUTIROSIN FROM THE ORGANISM.
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC   layer with hexagonal symmetry.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; L20421; AAA62588.1; -.
DR   PIR; T17884; T17884.
DR   InterPro; IPR003343; Big_2.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF02368; Big_2; 1.
DR   Pfam; PF00395; SLH; 1.
DR   SMART; SM00635; BID_2; 1.
KW   Signal; Cell wall; S-layer.
FT   SIGNAL   1   30   POTENTIAL.
FT   CHAIN   31   1616   S-LAYER RELATED PROTEIN.
SQ   SEQUENCE   1616 AA; 172874 MW; 1F03FF7A4F113AA7 CRC64;

Query Match   4.7%; Score 124; DB 1; Length 1616;
Best Local Similarity 20.1%; Pred. No. 3.5;
Matches 120; Conservative 78; Mismatches 214; Indels 184; Gaps 31;

Qy   1 MRKLNSLFSLAVLLSLL--CWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRD---P 55
            :::.|:| .||: :::. | ||:. ||| ::. | ::: | |||.|: |
Db   5 LRKWNGMMIIALVISLLTPAWGKASAQSLGQESGGV---QPQSAGVTDDVYVLSRDGTFR 61

Qy   56 IPAGMARIILEAHDVWEDGTGYQMLWDADHNQYGASI---PEBSFWFANGTIPAGLYDPF 112
            | ::: .| |..|||| | | :|| | .|: |::|||:|||| ||| |
Db   62 LPVGHINVNVDAP--PQKTNYVALFTS----GAQVTNSQETDKVFVKKTNTAIAVDKH 113

Qy   113 EYKVPVNADASFSPTNFVLDGTASADIPAGTY-----------------------DVV 147
            :|| | || |: |||| || || ||||
Db   114 DQVIRVIGPTAVPPTGSTWEENQNLPIPEGGYVLLANDSSWGTSVTRKPLFEHYKTGDTV 173

Qy   148 -----------IINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFTVQRQGPGDAASVVV 194
                       ||| |. ||: |.: : .|:| :| ||| ::|:: ||.|.| :
Db   174 SLHKGGQVVHAGDFLNPDPGLNLVTASGTT----VTSPDFTVSGQVVRYGSGGQISLTV 228

Qy   195 TG-----------EGGNEFAPVQNLQ--WSVSGQTVLTW----QAPASDKRTYV 232
            || || ||| | || |:|:|.|| | .|| .||||
Db   229 NGTEAALKADGAFQSAVRLTAGTNAISVKLLKDGREIVSSTVTVTYNDAQQPA---- 281

Qy   233 LNESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLS 292
            | .. :.||||..|: ::| : |: | | . | ||| || : :| .|..|:|
Db   282 --DLIEVEAAPIDIT-ISIEGPAHA-IGYVD-QDIAGIDDTVALFTNDW--------- 325

Qy   293 PDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAAANFTIKLLEETLGSDK 352
            | :|:||||:||| ||| :|| ||:||||| | | || | ::|| |: ||:|:|
Db   326 ------GPQITVPQ------FNVAVQV-------------DAGSKVTKVVNPSIDGKT 358

Qy   353 PA---PMNL-VKSEGVKLPAPYQERTIDLSAYAGQQV-YLA----------------- 388
            || ::| |:.|||| | | :: | | :||||| :|:
Db   359 PAWTGPTDLEIPSGGYVLVA--QD-----TSYAGKNIKKYLATYFKVGDAIKLRKNGFAV 411

Qy   389 -FRHFNSTG--IFRLYLDDVAVSGEGS-SNDYTYTVYR----DNVIAQNLAATTFNQEN 440
            . ||:|:.| || | .| :|.| :.||| ||:|| || ||| | :|:||
Db   412 PVKDLMGTGGPIARVTLDNYAMYTETKPSTELSGTITNMDDPSKIALTVNGTPLPFGPDG 471

Qy   441 VAPGQYNVCVEVKYTAGVSPKVCKDVTVEGSNEFAH---VQNLTGSAVGQKVTLKW 493
            | || |||| | || ||:||:|| |||||| :||| | |::| | :|| |
Db   472 KFKTSYTLAEGINYLDLV-------VTKEGKEQDSKDLVVYSRPGFSTGKKVIL-W 519

RESULT 13
SLAP_BACCI
ID   SLAP_BACCI   STANDARD;   PRT;   1616 AA.
AC   P35824;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   S-layer related protein precursor.
GN   BUTB.
OS   Bacillus circulans.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1397;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL B-3312;
RX   MEDLINE=94374689; PubMed=7522196;
RA   Aubert-Pivert E., Davies J.;
RT   "Biosynthesis of butirosin in Bacillus circulans NRRL B3312:
RT   identification by sequence analysis and insertional mutagenesis of
RT   the butB gene involved in antibiotic production.";

DR   EMBL; X92982; CAA63564.1; -.
DR   InterPro; IPR002479; CW_binding.
DR   InterPro; IPR007577; Gly_transf_sug.
DR   Pfam; PF01473; CW_binding_1; 28.
DR   Pfam; PF04488; Gly_transf_sug; 1.
KW   Toxin; Enterotoxin.
SQ   SEQUENCE   2710 AA; 308052 MW; 0A6E52CEB4C14421 CRC64;

Query Match   4.8%; Score 127; DB 1; Length 2710;
Best Local Similarity 20.4%; Pred. No. 4.4;
Matches 114; Conservative 66; Mismatches 200; Indels 178; Gaps 31;

Qy   72 EDGTGYQMLWDADHNQYGAS---------IPEBSFWF-ANGTIPAGLY--DPFEYK 115
            | |:   .| : |. : . |::| | | ::| |||: || ||
Db   2117 EAATGWQTI--DGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYF 2173

Qy   116 VPVNADAS------FSPTNFVLDGTA--SADIPAGTYDYVI-----INPNPGI--- 155
            .|.| |. || ||| || . |||| ||| :.   :| |.||
Db   2174 APANTDANNIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIILNNKKKYFNPNNAIAAI 2233

Qy   156 ------IYIVGEGVSKGNDYVVEAGKTYHFTVQRQGPGDAASVVVTG----EGGNE 201
                  |||| |||| ||:| | | ||:|:: .   :|| :. || |||.|
Db   2234 HLCTINNDKYYFSYDGIIQ-NGYITIERNNFYFDANNE------SKMVTGVFKGPNGFEY 2286

Qy   202 FAPVQNLQWSVSGQTVT----LTWQAPASDKRTYVLNESFDTQTLPNGWTMIDADGDGH 256
            || |:||| :| |.| : | : |:|| ||:. ..| ||||.|| :|||.||
Db   2287 FAPANTHNNNIEGQAIVVQNKFLT----LNGKKYYFDNDS---KAVTGWQTI----DGK 2334

Qy   257 NWLSTINVYNTAT--HTGDGAMF------SKSWTASGGAKIDLSPDNYLVTPKVTVP 305
            :| | | |.|. .|||| | | | .| |::|. : |:|||:|:|
Db   2335 KYYFNLNTAEAATGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNTNTFIASTGYT-S 2393

Qy   306 ENGKLSYWVSSQVPWTNEHYGVFLSTTG---------NEAAN-----------FTIKLLE 345
            .||| |: ||::| |: | | :.|| ||| | ||: |
Db   2394 INGKHFYFNTDGI------------MQIGVFKGPNGFEYFAPANTDANNIEGQAILYQNKFLTLNGKK 2449

Qy   346 ETLGSDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQQ------------------V 385
            :|||:|| | |:|| | | |: | | | :|| ||:
Db   2450 YYFGSDSKAVTGLRTIDGKK----YYFNTNTAVAVTGWQTINGKKYYFNTNTSIASTGYT 2505

Qy   386 YLAFRH--FNSTGIFRLYL------------DDVAVSGEGSSNDYTYTVYRDNV-VI 427
            ||| || |:|.||||||:   .|.| | ||| :||| | |.|. ||
Db   2506 IISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQNRFLYLHDNIYYF 2565

Qy   428 AQNLAATT----------FNQENVAPGQVNY-CVEVK--YTAGVSPKVCKDVTVEGSNEF 474
            | |||.|| ||::|.||| :| |:. | ::| : |:|| |||::| .|
Db   2566 GNNSKAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQI--GVFKGSNGF 2622

Qy   475 AH------VQNLTGSAV 485
            || .||:|||.|
Db   2623 EYFAPANTDANNIEGQAI 2640

RESULT 14

```
CTA1_BACCI
ID   CTA1_BACCI     STANDARD;      PRT;   972 AA.
AC   P94286;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Cycloisomaltooligosaccharide glucanotransferase precursor (EC 2.4.1.-)
DE   (CITASE).
OS   Bacillus circulans.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1397;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T-3040;
RA   Oguma T., Kurokawa T., Tobe K., Kobayashi M.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PRODUCES CYCLOISOMALTOOLIGOSACCHARIDE FROM DEXTRAN.
CC   -!- SIMILARITY: Belongs to family 66 of glycosyl hydrolases.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; D61382; BAA09604.1; -.
DR   InterPro; IPR005084; CBM_6.
DR   InterPro; IPR008979; Gal_bind_like.
DR   Pfam; PF03422; CBM_6; 2.
KW   Transferase; Glycosyltransferase; Signal.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39    972       CYCLOISOMALTOOLIGOSACCHARIDE
FT                                GLUCANOTRANSFERASE.
SQ   SEQUENCE   972 AA;  107431 MW;  2827BEA61782CC22 CRC64;

Query Match    4.6%;
Best Local Similarity  20.8%;
Matches   71;  Conservative  51;  Mismatches  111;  Indels  109;  Gaps  17;

Qy    183  RQGPGDAASVVVTGEGGN---------EFAPVQNLQWSVS------GQ--TVRLTWQA  223
           |::|: |       |  |        :||              ||  |||  |
Db     53  RYNPGDAVSIRVQAKNGTGSSWSGAARLEIFHLENSVYTSSQSLSLTNGQSTTLTFTWTA  112

Qy    224  PASDKRTYVLNESFDTQTLPNGWTMIDADGD-----GHNWLSTINVYNTATHTGDGAMFS  278
           |::|:|      ||:|    ::        :  ::|: |:|  |:  ::
Db    113  PSTDFRGYFVR--IDAGTLGQGATAIDVSSDFTKYPRYGYISEF------ESGETALES  163

Qy    279  KSWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAAN  338
           |      :|:::   |       |:|  ::   :  ||:   :||  ||
Db    164  K-------AKVDQLAQDY----------------HINAWQFYDWMWRHD----------  189

Qy    339  FTIKLLEETLGSDKPAPMNIVKSEGVKLPAPYQERTIDL------SAYAGQQVYLAFRHF  392
              ::|: | :   |                 |   :||     |      | :||| |
Db    190  --KMIKRTGGSIDSTWLDLFNRE--ISWSTLQNQIDAVHDVNGKAMAYAMIYASRENY  243

Qy    393  NSTGI---FRLYLDDVAVS-----GEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAP  443
           |:|:     |  |   |:|        ::|   |  |:  | :|||  |: | ||:|
Db    244  SPLGLSPTWGIYEDSSHTNQFDVDFGDGS---TYLYMSD----PQN----------P  283

Qy    444  GQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAV  485
           | | | ||  |:|   |: | ||:|     | ||| | :|
Db    284  NWQNY-IHAEYIDSINTAGFDGIHVDQMGQRSNVYDYNGNSI  324

RESULT 15
PICP_PSESR
ID   PICP_PSESR     STANDARD;      PRT;   587 AA.
AC   P42790;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Pseudomonalisin precursor (EC 3.4.21.100) (Pepstatin-insensitive
DE   carboxyl proteinase) (Pseudomonapepsin).
GN   PCP.
OS   Pseudomonas sp. (strain 101) (Achromobacter parvulus T1).
OC   Bacteria; Proteobacteria.
OX   NCBI_TaxID=33067;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 216-224.
RX   MEDLINE=95014496; PubMed=7929375;
RA   Oda K., Takahashi T., Tokuda Y., Shibano Y., Takahashi S.;
RT   "Cloning, nucleotide sequence, and expression of an isovaleryl
RT   pepstatin-insensitive carboxyl proteinase gene from Pseudomonas sp.
RT   101.";
RL   J. Biol. Chem. 269:26518-26524 (1994).
RN   [2]
RP   SEQUENCE OF 216-585.
RX   MEDLINE=96157710; PubMed=8576087;
RA   Hayashi K., Izu H., Oda K., Fukuhara K.-I., Matsuo M., Takano R.,
RA   Hara S.;
RT   "The primary structure of pepstatin-insensitive carboxyl proteinase
RT   produced by Pseudomonas sp. No. 101.";
RL   J. Biochem. 118:738-744(1995).
RN   [3]
RP   MUTAGENESIS OF ASPARTIC ACID AND GLUTAMIC ACID RESIDUES, AND ACTIVE
RP   SITES.
RX   MEDLINE=99419069; PubMed=10488127;
RA   Oyama H., Abe S.-I., Ushiyama S., Takahashi S., Oda K.;
RT   "Identification of catalytic residues of pepstatin-insensitive
RT   carboxyl proteinases from prokaryotes by site-directed mutagenesis.";
RL   J. Biol. Chem. 274:27815-27822(1999).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 216-585.
RX   MEDLINE=21614391; PubMed=11747435;
RA   Wlodawer A., Li M., Gustchina A., Dauter Z., Uchida K., Oyama H.,
RA   Goldfarb N.E., Dunn B.M., Oda K.;
RT   "Inhibitor complexes of the Pseudomonas serine-carboxyl proteinase.";
RL   Biochemistry 40:15602-15611(2001).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.0 ANGSTROMS) OF 216-587.
RX   MEDLINE=21223584; PubMed=11323721;
RA   Wlodawer A., Li M., Dauter Z., Gustchina A., Uchida K., Oyama H.,
RA   Dunn B.M., Oda K.;
RT   "Carboxyl proteinase from Pseudomonas defines a novel family of
RT   subtilisin-like enzymes.";
RL   Nat. Struct. Biol. 8:442-446(2001).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of the B chain of insulin at 13-
CC       Glu-|-Ala-14, 15-Leu-|-Tyr-16 and 25-Phe-|-Tyr-26 and angiotensin
CC       I at 4-Tyr-|-Ile-5. A good synthetic substrate is Lys-Pro-Ile-Glu-
CC       Phe-|-Phe (NO(2)|-Arg-Leu.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit.
CC   -!- SUBCELLULAR LOCATION: Periplasmic.
CC   -!- PTM: Autocatalytically processed.
CC   -!- SIMILARITY: Belongs to peptidase family S53.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; D37970; BAA07188.1; -.
DR   PDB; 1GA1; 11-FEB-03.
DR   PDB; 1GA4; 11-FEB-03.
DR   PDB; 1GA6; 11-FEB-03.
DR   PDB; 1KDV; 11-FEB-03.
DR   PDB; 1KDY; 11-FEB-03.
DR   PDB; 1KDZ; 05-FEB-03.
DR   PDB; 1KE1; 11-FEB-03.
DR   PDB; 1KE2; 11-FEB-03.
DR   MEROPS; S53.001; -.
DR   Pfam; PF00082; Peptidase_S8; 1.
```

```
KW   Hydrolase; Protease; Serine protease; Zymogen; Periplasmic; Signal;
KW   3D-structure; Calcium-binding;
FT   SIGNAL         1      ?          POTENTIAL.
FT   PROPEP         ?     215         REMOVED IN MATURE FORM.
FT   CHAIN        216     585         PSEUDOMONALISIN.
FT   PROPEP       586     587         REMOVED IN MATURE FORM.
FT   ACT_SITE     295     295         CHARGE RELAY SYSTEM.
FT   ACT_SITE     299     299         CHARGE RELAY SYSTEM.
FT   ACT_SITE     502     502         CHARGE RELAY SYSTEM.
FT   DISULFID     352     391
FT   METAL        543     543         CALCIUM.
FT   METAL        544     544         CALCIUM (VIA CARBONYL OXYGEN).
FT   METAL        559     559         CALCIUM (VIA CARBONYL OXYGEN).
FT   METAL        561     561         CALCIUM (VIA CARBONYL OXYGEN).
FT   METAL        563     563         CALCIUM.
SQ   SEQUENCE     587 AA;   61072 MW;  E193D8B2C225829A CRC64;

Query Match               4.6%;   Score 120.5;   DB 1;   Length 587;
Best Local Similarity    23.5%;   Pred. No. 1.6;
Matches  78;   Conservative  33;   Mismatches  108;   Indels  113;   Gaps   16;

Qy    103  TIPAGL--YDPPEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIV  159
                  :  ::|  ||:   :|:
Db    214  TLAAGTAKGHNPTEF--PTIYDASSAPT-------------------AANTVGIITIG  251

Qy    160  G-------------EGVSKGNDYVVEAGKT---YHFTVQRQGPGDAASVVVTGEGGNEF  202
                         | ::         :    :::
Db    252  GVSQTLQDLQQFTSANGLASVNTQTIQTGSSNGDYSDDQQGQGEWDLDSQSIVGSSAG---  308

Qy    203  APVQNLQW-----SVSGQT-VTLTWQAPASDKRTYVLNESFDTQTLPNGWTMIDADGDGH  256
                 | ||||        ||:|:             |:|            ||
Db    309  GAVQQLLFYMADQSASGNTGLTQAFNQAVSDNVAKVINVSL-----GWCEADANADG-  360

Qy    257  NWLSTINVYNTATHTG------------------DGAMFSKSWTASGGAKIDLSPDN  295
                  ::                            ||:||||   ||:
Db    361  TLQAEDRIFATAAAQGGTFSVSSGDEGVYECNNRGYPDGSTYSVSWPAS------SPNV  413

Qy    296  YLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLG----SD  351
                  ::  |          |:|        :  ::|
Db    414  IAV-------GGTILYTTSAGAYSNE------TVWNEGLDSNGKLWATGGGYSVYES  457

Qy    352  KPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQ  383
                 ||:  |||  :  |:|  |:|
Db    458  KPSWQSVVSG-----TPGRRLLPDISFDAAQ  483


Search completed: May 18, 2004, 11:43:30
Job time : 10.2836 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: May 18, 2004, 11:36:20 ; Search time 36.3342 Seconds
(without alignments)
4315.838 Million cell updates/sec

Title: US-08-570-311-2
Perfect score: 2641
Sequence: 1 MRKLNSLFSLAVLLSLLCWG.........QNLTGSAVGQKVTLKWDAPN 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB set length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	951.5	36.0	1706	2	Q51839	porphyromon
2	950.5	36.0	1706	2	Q51838	porphyromon
3	944.5	35.8	1704	2	Q51816	porphyromon
4	943.5	35.7	1687	2	Q9R9B7	porphyromon
5	824	31.2	1723	2	P72194	porphyromon
6	818	31.0	1723	2	P72197	porphyromon
7	816	30.9	1358	2	P96967	porphyromon
8	812.5	30.8	1732	2	Q51817	porphyromon
9	807.5	30.6	1732	2	O52050	porphyromon
10	806.5	30.5	1732	2	O07442	porphyromon
11	805.5	30.5	1223	2	Q92NB5	porphyromon
12	561.5	21.3	1097	2	P72196	porphyromon
13	379	14.4	312	2	Q9KIB3	porphyromon
14	337	12.8	925	2	Q9F4J0	porphyromon
15	213.5	8.1	293	2	Q9XBU9	porphyromon
16	156	5.9	1454	16	Q8ENZ2	oceanobacil
17	154	5.8	1155	17	Q8TPZ8	methanosarc
18	153	5.8	5291	16	Q8X2T1	escherichia
19	152	5.8	1234	2	Q60046	thermoanaer
20	150	5.7	1483	16	Q977Y4	clostridium
21	147	5.6	1983	16	Q9CJ55	lactococcus
22	146	5.5	2951	17	Q8THC9	methanosarc
23	144.5	5.5	781	16	Q9S2G1	streptomyce
24	139	5.3	5020	16	Q8E9W3	shewanella
25	138.5	5.2	2468	16	Q9I2M3	pseudomonas
26	137.5	5.2	8173	16	Q7UDU8	rhodopirell
27	137	5.2	2013	16	Q92EK2	listeria in
28	135	5.1	1615	2	Q9F0P9	rickettsia
29	135	5.1	1654	2	Q93QW9	rickettsia
30	135	5.1	3064	16	Q82XT8	nitrosomona
31	133	5.0	897	16	Q8G7K7	bifidobacte
32	133	5.0	1734	17	Q8TI67	methanosarc
33	132	5.0	698	16	Q82F44	streptomyce
34	131.5	5.0	706	16	Q93J50	streptomyce
35	131.5	5.0	768	16	Q88G76	pseudomonas
36	131.5	5.0	1379	17	Q9HL55	thermoplasm
37	131.5	5.0	1618	2	Q9KKB4	rickettsia
38	131	5.0	1357	17	Q8THC8	methanosarc
39	131	5.0	1618	2	Q9KKB1	rickettsia
40	131	5.0	2009	16	Q7UY44	rhodopirell
41	131	5.0	2768	16	Q8B9G6	shewanella
42	130.5	4.9	524	16	Q7UUW6	rhodopirell
43	130.5	4.9	1616	2	Q9KKB0	rickettsia
44	130.5	4.9	1617	2	Q9KKB9	rickettsia
45	130.5	4.9	2217	17	Q8TP72	methanosarc

ALIGNMENTS

RESULT 1
Q51839
ID   Q51839   PRELIMINARY;   PRT;   1706 AA.
AC   Q51839; Q51840;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arginine-specific thiol protease precursor.
GN   PRTR.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=95160709; PubMed=7857299;
RA   Kirszbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N.,
RA   Reynolds E.C.;
RT   "Complete nucleotide sequence of a gene prtR of Porphyromonas
RT   gingivalis W50 encoding a 132 kDa protein that contains an arginine-
RT   specific thiol endopeptidase domain and a haemagglutinin domain.";
RL   Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96311339; PubMed=8713096;
RA   Slakeski N., Cleal S.M., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtR that encodes
RT   an arginine-specific thiol proteinase and multiple adhesins.";
RL   Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Reynolds E.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;

```
Db   1323  YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN 1381

Qy    476  HVQNLTGSAVGQKVTLKWDAPN 497
           |||   :  ||:|:|:
Db   1382  PVKNLKAQPDGGDVVLKWEAPS 1403
```

RESULT 2
Q51838

```
ID   Q51838          PRELIMINARY;        PRT;  1706 AA.
AC   Q51838;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Protease precursor.
GN   PRPR1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96071894; PubMed=7591131;
RA   Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.;
RT   "Characterization, genetic analysis, and expression of a protease
RT   antigen (PrpRI) of Porphyromonas gingivalis W50.";
RL   Infect. Immun. 63:4744-4754(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.;
RT   "The prpR1 and the prR2 arginine-specific protease genes of
RT   Porphyromonas gingivalis W50 produce five biochemically distinct
RT   enzymes.";
RL   Mol. Microbiol. 23:0-0(1997).
DR   EMBL; X82680; CAA57997.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN       228    719       ALPHA-PROTEASE.
FT   CHAIN       720   1262       BETA-ADHESIN.
SQ   SEQUENCE   1706 AA;  185705 MW;  0E56DCD87FDA8CDD  CRC64;
```

Query Match
Best Local Similarity   36.0%;   Score 950.5; DB 2; Length 1706;
                        31.8%;   Pred. No. 3.1e-50;
Matches 236;   Conservative   77;   Mismatches   154;   Indels 275; Gaps   20;

```
Qy     22  TAAAQG-------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH  68
           |||          || :| ||:| |:  |:|      | |: ||  ||||  |||||||
Db    671  TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy     69  DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
           | |:|| ||||| |||||| || ||||    |||| ||||  :  | ||| |||| |:
Db    731  DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

Qy    127  TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
           ||| |:| || |||||||| ||||  ||   | ||  |  ||||  ||| ||:| :  :
Db    790  TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGGQPTKEDDYVFEAGKKYHFLMKKMG 849

Qy    186  PGDAASVV----------------------------------------------------- 193
           |  ||||
Db    850  SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 909

Qy    194  ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------------DKRTYVLNE 235
                 | | || | || ||:|  |:  :  :|   |            ||| ||:||
Db    910  PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNTPNPNPNPNPGTTTLSE 969

Qy    236  SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
           :  :   ||  :|  |   |     |  | |: :| |     |:   | | |:||:|||
Db    970  SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1025

Qy    296  YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETL---GSD 351
           ||:||  ||  ||| :  | :    |  : ||| |:|| |:|| :| |  ||||    |
Db   1026  YLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1085

Qy    352  KPAPM------------------------------------------------------- 356
           || ||
Db   1086  SPEAMRGRIQGTWRQKTVDLPAGTKVYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1145

Qy    357  -----------------------------------------NLVKS-------------- 361
                                                    ||:||
Db   1146  ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALNPDNYL 1205

Qy    362  ---------------------------------------------------------- 361
Db   1206  ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1265

Qy    362  ---EGVKLPAPYQERRIDLSAYAGQQVYLARFRHFNSTGIFRLYLDDV--AVSGEGSSND 415
              | |||||  |:| ||||||| ||:| |:||||||:||| ||| ||     || | |
Db   1266  STEADGAKPQSVWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD 1322

Qy    416  YTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA 475
           |||||||| ||:|:|| || |::|| |||:||||||||||||||||| |||||||||||
Db   1323  ...
```

```
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0005508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
FT   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   CHAIN    228    719    MATURE 50-KDA CYSTEINE PROTEINASE
FT                         GINGIPAIN.
SQ   SEQUENCE   1704 AA;  185436 MW;  6A34B40131C2A676 CRC64;


Query Match    35.8%;  Score 944.5;  DB 2;  Length 1704;
Best Local Similarity  31.9%;  Pred. No. 7.4e-50;
Matches  235;  Conservative  75;  Mismatches  160;  Indels  267;  Gaps  20;


Qy    22  TAAAQG--------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH    68
          ||||                                          : :: :||||||
Db   671  TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH   730

Qy    69  DVWEDGTGYQMLWDADHNQYGASIPEE--SFWANGTIPAGLYDPFFEYKVPVNADASFSP   126
          |||::||:||:|:||||:||| :||:       | ::|:   :||||:||:||||:|:||
Db   731  DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP   789

Qy   127  TNFVLDGTASADIPAGTYDYVLINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG   185
          ||.::||||| :|||||||:.:: .| :  | ::|:   .:|||:.|||.|||| : :.|
Db   790  TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG   849

Qy   186  PGDAASVV-                                                     193
          .||:
Db   850  SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS   909

Qy   194  ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS----------DKRTYVLNESF   237
                |||.|.||||||||||  .|:.|:||||.||:|           |||||:|.||
Db   910  PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTLSESF    969

Qy   238  DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL   297
           : : |||:|:|||||||| | | :.::|   |. |.:    ||:.|  |:|:|:.|||:
Db   970  E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL  1025

Qy   298  VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETLGS------   350
          :||:: :|:||||::||  | ||:.:|||.|..::.|:::|||  ||||||    :
Db  1026  ITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy   351  ---------------DKPAPMNLV------------KSEGVKL                   366
                         |||
Db  1086  EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145

Qy   367  ----PAPYQERTID--------------LSAYAGQQV----                     385
              |||
Db  1146  STHGEARPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy   386  ------                                                         385
Db  1206  KDVTGATKVKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST 1265

Qy   386  ------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV                   420
Db  1266  EANGAKPQSVWIERTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES  1325

Qy   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
          |||:. ||| ||||||:|||:|||:|||||||||||||||: |:||.|:.:||||:|:|:
Db  1326  YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1384

Qy   481  TGSAVGQKVTLKWDAPN                                             497
          .|.:.|:.|:|||.||:
Db  1385  KAQPDGGDVVLKWRAPS                                            1401


RESULT 4
Q9R9B7
ID   Q9R9B7                  PRELIMINARY;       PRT;   1687 AA.
```

```
Db   790  TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG  849

Qy   186  PGDAASVV-                                                     193
          .||:
Db   850  SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVS  909

Qy   194  ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS----------DKRTYVLNE    235
                |||.|.||||||||||  .|:.|:||||.||:|           ||||||:|.
Db   910  PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTLSE     969

Qy   236  SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN  295
          ||: : |||:|:|||||||| | | :.::|   |. |.:    ||:.|  |:|:|:.|||
Db   970  SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1025

Qy   296  YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS----  350
          ||:||:: :|:||||::||  | ||:.:|||.|..::.|:::|||  ||||||
Db  1026  YLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1085

Qy   351  ----------------DKPA------------------                        354
                          |||
Db  1086  SPEAIRGRIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF  1145

Qy   355  ------PMNLV                                                    359
Db  1146  ESSTHGEATAEWTTIDADGDGCQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALNPDNYL 1205

Qy   360  ------                                                         359
Db  1206  ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL  1265

Qy   360  --KSEGVKLPAPYQERTIDLSAYAGQQVILAPRHFNSTGIFRLYLDDV--AVSGEGSSND  415
            :||.|||.|.||||| : |||.|.||||: ||||:|||||||||::|   .|.|.||:|
Db  1266  STEADGAKPQSVWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD 1322

Qy   416  YTYTYVRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA  475
          ||||||||:. ||| ||||||:|||:|||:|||||||||||||||: |:||.|:.:|||:
Db  1323  YTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN 1381

Qy   476  HVQNLTGSAVGQKVTLKWDAPN                                         497
          :.|:.|.:.|:.|:|||.||:
Db  1382  PVKNLKAQPDGGGDVVLKWRAPS                                       1403


RESULT 3
Q51816
ID   Q51816                  PRELIMINARY;       PRT;   1704 AA.
AC   Q51816;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arg-gingipain-1 proteinase.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95138080; PubMed=7836351;
RA   Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
RA   Travis J., Barr P.J.;
RT   "Molecular cloning and structural characterization of the Arg-
RT   gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a
RT   proteinase-adhesin polyprotein.";
RL   J. Biol. Chem. 270:1007-1010(1995).
DR   EMBL; U15282; AAA69539.1; -.
DR   PIR; A55426; A55426.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
```

```
AC   Q9R9B7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin/protease.
GN   HAGE.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Dong H., Progulske-Fox A.;
RT   "Cloning and characterization of hagE from P. gingivalis 381.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF026946; AAD01810.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE   1687 AA;  183702 MW;  D085B16A399FE70 CRC64;

Query Match             35.7%;  Score 943.5;  DB 2;  Length 1687;
Best Local Similarity   31.8%;  Pred. No. 8.4e-50;
  Matches  235;  Conservative   75;  Mismatches  160;  Indels  269;  Gaps  20;

Qy         22 TAAAQG----------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
                 |||||           : ||  :| ||  ||   :       |||| ||  ||||   |||:|||
Db        652 TATTQQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 711

Qy         69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
                 |||:|||:|||||||||||   :  ||    ||||||||||||||  | |||||||||||||
Db        712 DVWNDGSGYQILLDADHDQYGQVIPSDTHTIM--PNCSVPANLFAPFEYTVPENADPSCSP 770

Qy        127 TNFVLDGTASADIPAGTYDVVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
                 ||||||||| |||||||||||  |||||:|  |||||||| ||||||| |  ||||||: :
Db        771 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVPEAGKKYHFLMKKMG 830

Qy        186 PGDAASVV--------------------------------------------- 193
                 |||: |
Db        831 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 890

Qy        194 -----VTGEGSNEFAPVQNLQWSVSGQTVTLTWQAPAS-------------DKRTYVLNE 235
                      |||:|||||||| | |:  : || |||   |||             |||| ||:
Db        891 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNPNPNPNPNPNPGTTLSE 950

Qy        236 SFDTQTLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
                  |: : |||| ||||||| | |: ||:|:| |  ||| | |: ||    || ||  |
Db        951 SFE-NGIPASNKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1006

Qy        296 YLVTEKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
                  |:| |::|||||| | |:   :   : : :||| |  |||   |||| ||  || |
Db       1007 YLIITPALDLPNGGKLTFWVCAQDANYASHYAVVASSTGNDASNFTNALLEETITAKGVR 1066

Qy        351 ------------DKPAPMNLV--------------------------KSEGVKL------ 366
                            |   || :|                          ||||:|
Db       1067 SPEAIRGRIGTGTWRQKVTDLLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1126

Qy        367 ------PAPYQERTID-------------------------------LSAYAGQQV---- 385
```

```
Db       1127 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL 1186
                   |:|:   :  |                                   |:|:|   ||
Qy        386 --------------------------------------------- 385

Db       1187 ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVPEETNGINKGGARFGL 1246
                                :|||||||    :   |    :  |    |        :|    :
Qy        386 -------------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTY 418

Db       1247 STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPFPTDYTY 1306
                                              |:||||||::   :|:||||||||| ||||
Qy        419 TVYRDNVVIAQNLAATFNQENVAPGQYNYCVEVKYTAGVSPXVCKDVTVEGSNEFAHVQ 478
                 |||||  :  |  :|  ||| |||||:| ||||||||||| |  | ::
Db       1307 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVK 1365

Qy        479 NLTGSAVGQKVTLKWDAPN 497
                 ||   | ||:|:|||:||
Db       1366 NLKAQPDGGDVVLKWEAPS 1384


RESULT 5
P72194
ID   P72194      PRELIMINARY;      PRT;   1723 AA.
AC   P72194;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97044756; PubMed=8889827;
RA   Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
RT   "Cloning and sequencing of the gene encoding a novel lysine-specific
RT   cysteine proteinase (Lys-gingipain) in Porphyromonas gingivalis:
RT   structural relationship with the arginine-specific cysteine proteinase
RT   (Arg-gingipain).";
RL   J. Biochem. 120:398-406(1996).
DR   EMBL; D83258; BAA11870.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1723 AA;  187261 MW;  5628963D251493EB CRC64;

Query Match             31.2%;  Score 824;  DB 2;  Length 1723;
Best Local Similarity   29.8%;  Pred. No. 2.3e-42;
  Matches  220;  Conservative   75;  Mismatches  170;  Indels  274;  Gaps  21;

Qy         22 TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD---------PIPAGMARIILEA 67
                 |||   ||       |  ::  |  : ::                 |::|:
Db        693 TATTQQKVTLKWDAPS-----AKKAEASREVKRIGDGLFVTIEPANDVRANEAKVLAA 747

Qy         68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
                 |:|| | ||||| |||||:: ||:|||    |      ||| |  ||| :| :||||: |
Db        748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVTT 806

Qy        127 TNFVLDGTASADIPAGTYDVVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
```

```
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00069; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1723 AA;  186831 MW;  4508A7E50197CBBD CRC64;

Query Match             31.0%;  Score 818;  DB 2;  Length 1723;
Best Local Similarity   29.4%;  Pred. No. 5.4e-42;
Matches  218;  Conservative  78;  Mismatches 166;  Indels 280;  Gaps  21;

Qy      22 TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARILEA   67
Db     693 TATTQGQKVTLKWDAPS----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA  747

Qy      68 HDVWEDGTGYQMLWDADHNQYGASIPESFWFANGTIPAGLYDP-FEYKVPVNADASFSP  126
Db     748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVTT  806

Qy     127 TNFVLDGTASADIPAGTYDYVIINPNP---GIIYIVGEG---VSKGNDYVVEAGKTYHFTV  181
Db     807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM  866

Qy     182 QRQGPGDAASVV------------  193
Db     867 RRAGMGDGTDMEVEDDSPASYTYTVIRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT  926

Qy     194 -----------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTYVLNE  235
Db     927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSE  986

Qy     236 SFDTQTLPNGWTMIDADGGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN  295
Db     987 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1042

Qy     296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNRAANFTIKLLEETLGS----  350
Db    1043 YLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1102

Qy     351 ----DKPA------------------------------  354
Db    1103 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1162

Qy     355 ----------------------------PMNLV  359
Db    1163 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL 1222

Qy     360 --------------------------------------------  359
Db    1223 ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1282

Qy     360 --KSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDV--AVSGEGSSND  415
Db    1283 STEADGAKPQSVWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD 1339

Qy     416 YTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA  475
Db    1340 YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTI-NPTQFN 1398

Qy     476 HVQNLTGSAVGQKVTLKWDAPN 497
Db    1399 PVKNLKAQPDGGDVVLKWEAPS 1420
```

RESULT 7
P96967
```
ID   P96967         PRELIMINARY;       PRT;  1358 AA.
AC   P96967;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin.
```

```
Db     807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM  866

Qy     182 QRQGPGDAASVV------------  193
Db     867 RRAGMGDGTDMEVEDDSPASYTYTVIRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT  926

Qy     194 -----------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTYVLNE  235
Db     927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSE  986

Qy     236 SFDTQTLPNGWTMIDADGGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN  295
Db     987 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1042

Qy     296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNRAANFTIKLLEETLGS----  350
Db    1043 YLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1102

Qy     351 ----PAPYQERTID------------------LSAYAGQQV----  366
Db    1103 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1162

Qy     367 -----PAPYQERTID------------------LSAYAGQQV-----  385
Db    1163 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL 1222

Qy     386 ------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTY  418
Db    1223 ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1282

Qy     386 --------------------------------------------------  385
Db    1283 STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTY 1342

Qy     419 TVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ  478
Db    1343 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCKCVNVTI-NPTQFNPVK 1401

Qy     479 NLTGSAVGQKVTLKWDAPN 497
Db    1402 NLKAQPDGGDVVLKWEAPS 1420
```

RESULT 6
P72197
```
ID   P72197         PRELIMINARY;       PRT;  1723 AA.
AC   P72197;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA   Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT   "Molecular cloning and characterization of Porphyromonas gingivalis
RT   Lys-gingipain.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U54691; AAA99810.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
```

GN   HAGD.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Lepine G., Whitlock J., Wojciechowski L., Progulske-Fox A.;
RT   "Cloning, sequencing and characterization of hagD, a member of the
RT   HArep multigene family in Porphyromonas gingivalis.";
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U68468; AAB49691.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   SMART; SM00060; FN3; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1358 AA; 147102 MW; 47FCA0B25B06DED8 CRC64;

Query Match   30.9%; Score 816; DB 2; Length 1358;
Best Local Similarity   29.6%; Pred. No. 5.1e-42;
Matches 219; Conservative   75; Mismatches 171; Indels 274; Gaps   21;

Qy   22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARILEA 67
            ||| |||| || ||| :|:||:||:| | :|:|||||||
Db   328 TATTQGQKVTLKWDAPS-----AKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAA 382

Qy   68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
           :||:|| ||||| |||||||||||||||||||||||||||||| |||:|||||||||||
Db   383 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 441

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
            || :| ::||||:| : ||| | || ::|:||||||||||||||||||||
Db   442 QNIIVTGQGEVVIPGGVVDYCIINPEPASGKMWIAGDGDNQPARYDDFTFEAGKKYTFTM 501

Qy   182 QRQGPGDAASVV------- 193
           :|:: ::::
Db   502 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 561

Qy   194 -----------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS------DKRTYVLNE 235
                      | || :|| || ::|||: :: :: |:
Db   562 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNPGTPNPNPNPGTTTLSE 621

Qy   236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
         |:   :|| || ||||||| | | | ::| | :   : :| : |::|:|||
Db   622 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 677

Qy   296 YLVTPKVTVPENGKLSTWVSSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETLGS---- 350
         ||:|| :: |:||| ||| :| : |:|||.| :|:||:|:||| .||||||
Db   678 YLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 737

Qy   351 ------------DKPAPMNLV-----------------KSEGVKL----- 366
                                                  :||
Db   738 SPEAIRGRIQGTWRQKYVDLPAGTKYTVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 797

Qy   367 -----PAPYQERTID---------LSAYAGQV----- 385
              ||||:   :
Db   798 ESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL 857

Qy   386 -- 385
Db   858 ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 917

Qy   386 --YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTY 418
                                                         :||||:| :| : :::
Db   918 STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTY 977

Qy   419 TVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ 478
         ||||||| |:| |||||||||||||||||||||||||||||||||:|::: |:|||||:|
Db   978 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNPVK 1036

Qy   479 NLTGSAVGQKVTLKWDAPN 497
         || |||||||||||||:|:
Db 1037 NLKAQPDGGDVVLKWEAPS 1055

RESULT 8
Q51817
ID   Q51817   PRELIMINARY;   PRT;   1732 AA.
AC   Q51817;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Porphypain.
GN   PRTP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W12;
RX   MEDLINE=96213011; PubMed=8631659;
RA   Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA   Progulske-Fox A., Lantz M.S.;
RT   "Analysis of the prtP gene encoding porphypain, a cysteine proteinase
RT   of Porphyromonas gingivalis.";
RL   J. Bacteriol. 178:2734-2741(1996).
DR   EMBL; U42210; AAB06565.1; -.
DR   PIR; T30836; T30836.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1732 AA; 187875 MW; 654271DBEF7BCAE4 CRC64;

Query Match   30.8%; Score 812.5; DB 2; Pred. No. 1.2e-41;
Best Local Similarity   30.0%; Mismatches 172; Length 1732;
Matches 222; Conservative   70; Indels 277; Gaps   22;

Qy   22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARILEA 67
            ||| |||| || ||| :|:||:||:| | :|:|||||||
Db   693 TATTQGQKVTLKWEAPS-----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy   68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
           :||:|| ||||| |||||||||||||||||||||||||||||| |||:|||||||||||
Db   748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 806

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
            || :| ::||||:| : ||| | || ::|:||||||||||||||||||||
Db   807 QNIIVTGQGEVVIPGGVVDYCIINPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 866

Qy   182 QRQGPGDAASVV 193
           :|:: ::::
Db   867 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

```
KW    Protease.
SQ    SEQUENCE   1732 AA;  187931 MW;  B2374463D5CB5EA5 CRC64;

Query Match             30.6%;  Score 807.5;  DB 2;  Length 1732;
Best Local Similarity   30.0%;  Pred. No. 2.4e-41;
Matches 222;  Conservative  69;  Mismatches 173;  Indels 277;  Gaps  22;

Qy     22 TAAAQQGPKT----APSVTHQAVQKGIRTSKVKDLRD------------PIPAGMARIILEA  67
             |||           :||                |:::|
Db    693 TATTQQGKVTLKWEAPS---AKKAEGSREVKRIGDLFVIEPANDVRANEAKVVLAA  747

Qy     68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADAGFSP  126
             ::::||||||              ::    :|||   :|   ::::|| :
Db    748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVTT  806

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVEGG---VSKGNDYVVEAGKTYHFTV  181
             :::      :|||||||||||||||||       ::::         :   ||||||||
Db    807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGNPARYDFTFEAGKKYTFTM  866

Qy    182 QRQGPGDAASVV--------------------------------------------  193
             :::
Db    867 RRAGMDGTDMEVEDDSPASYTYTVRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT  926

Qy    194 ----------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP----------ASDKRTYVLN  234
                       ||||||||||||||||||||||||||||||          :::
Db    927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS  986

Qy    235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPD  294
             |||: NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD
Db    987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD  1042

Qy    295 NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS-  350
             ||||| :: :::|   ||||||| |||||||||:|||||||||||||||||||||
Db   1043 NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV  1102

Qy    351 -------DKPAPMNLV----------------------------KSEGVKL-  366
                    |||||||||                            ||||||
Db   1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET  1162

Qy    367 ------PAPYQERTID------------------LSAYAGQQV------  385
                    |||||||                    ::|||
Db   1163 FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWITAHGGSNVVSSFSWNGMALNPDNY  1222

Qy    386 -------------------------------------------------------  385
Db   1223 LISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFG  1282

Qy    386 ---------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT  417
                                  :|||||||||||||||||||  ||||||||||
Db   1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNVILLDDIQFTMGGSPTPTDYT  1342

Qy    418 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV  477
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1343 YTVYRDGTNIKEGLTETTFEEYGVATGNHEYCVEVKYTAGVSPKKCVDVTV-NSTQFNPV  1401

Qy    478 QNLTGSAV--GQKVTLKWDAP  496
             |||||||  |||||||||||
Db   1402 QNLTAEQAPNSMDAILKWNAP  1422
```

```
RESULT 10
O07442
ID    O07442          PRELIMINARY;    PRT;  1732 AA.
AC    O07442;
DT    01-JUL-1997 (TrEMBLrel. 04, Created)
DT    01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE    Lysine-specific cysteine proteinase.
GN    PRTK.
OS    Porphyromonas gingivalis (Bacteroides gingivalis).
OC    Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC    Porphyromonadaceae; Porphyromonas.
OX    NCBI_TaxID=837;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=99235907; PubMed=10219167;
RA   Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtK that encodes
RT   a lysine-specific cysteine proteinase and three sequence-related
RT   adhesins.";
RL   Oral Microbiol. Immunol. 14:92-97(1999).
DR   EMBL; U75366; AAB60809.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS006697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1732 AA;   187914 MW;   45D5B91377391703 CRC64;

Query Match             30.5%;  Score 806.5;  DB 2;  Length 1732;
Best Local Similarity   29.7%;  Pred. No. 2.8e-41;
Matches  220;  Conservative  72;  Mismatches 172;  Indels 277;  Gaps  22;

Qy     22  TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD----------PIPAGMARIILEA 67
           |||   |           :: :|||: :|||:::|
Db    693  TATTQGGQKVTLKWEAPS----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy     68  HDVWEDGTGYQMLWDADHNQYGASIPESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
           :| ||||||||::||||||| ||||
Db    748  DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 806

Qy    127  TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
           ::
Db    807  QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 866

Qy    182  QRQGPGDAASVV------------ 193
           :|||
Db    867  RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy    194  ------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP----------ASDKRTYVLN 234
Db    927  AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNPNPNPNPNPGTTLS 986

Qy    235  ESFDTQTLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
Db    987  ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042

Qy    295  NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETLGS-- 350
Db   1043  NYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102

Qy    351  --------------DKPAPMNLV-------- 366
Db   1103  RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy    367  --------PAPYQERTID---- 385
Db   1163  FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy    386  -------------------------- 385
Db   1223  LISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTVVFETPNGINKGGARFG 1282

Qy    386  ------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
Db   1283  LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYT 1342
```

```
Qy    418  YTVVRDNVVIAQNLAATTENQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
           |||||:  |:|:||| :|||||||||||||||||||||||||:||||
Db   1343  YTVYRDGTKIKEGLTETTEEDGVATGNHEYCVEVKYTAGVSPKKCVNTV-NSTQFNPV 1401

Qy    478  QNLTGSAV--GQKVTLKWDAP 496
           ||:|| :      |||:||||
Db   1402  QNLTAEQAPNSMDAILKWNAP 1422
```

```
RESULT 11
Q92NB5
ID   Q92NB5            PRELIMINARY;    PRT;   1223 AA.
AC   Q92NB5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   130k-HMGD (Fragment).
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=99143166; PubMed=9988746;
RA   Shibata Y., Hayakawa M., Takiguchi H., Shiroza T., Abiko Y.;
RT   "Determination and characterization of the hemagglutinin-associated
RT   short motifs found in Porphyromonas gingivalis multiple gene
RT   products.";
RL   J. Biol. Chem. 274:5012-5020(1999).
DR   EMBL; AB019363; BAA34341.1; -
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS006697; DNA_LIGASE_A1; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   1223 AA;   131542 MW;   00225CD2BA9F91B3 CRC64;

Query Match             30.5%;  Score 805.5;  DB 2;  Length 1223;
Best Local Similarity   29.6%;  Pred. No. 2e-41;
Matches  219;  Conservative  73;  Mismatches 172;  Indels 277;  Gaps  22;

Qy     22  TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD----------PIPAGMARIILEA 67
           |||   |           :: :|||: :|||:::|
Db    184  TATTQGGQKVTLKWEAPS----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 238

Qy     68  HDVWEDGTGYQMLWDADHNQYGASIPESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
           :| ||||||||::||||||| ||||
Db    239  DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 297

Qy    127  TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
           ::
Db    298  QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 357

Qy    182  QRQGPGDAASVV------------ 193
           :|||
Db    358  RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 417

Qy    194  ------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP----------ASDKRTYVLN 234
Db    418  AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNPNPNPNPNPGTTLS 477

Qy    235  ESFDTQTLPNGWTMIDADGDGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
```

Query Match   21.3%;   Score 561.5;   DB 2;   Length 1097;
Best Local Similarity   29.8%;   Pred. No. 2.4e-26;
Matches 158;   Conservative 48;   Mismatches 117;   Indels 207;   Gaps 17;

```
Qy   165  KGNDYVVEAGKT--YHFT------VQRQGFPGDAASVVVTGEGGNEFAPVQNLQWSVSG  214
              ||| :  |  |||  :  :          |||||||||:||||||||:|
Db   267  KGSDYTVATGRLRFGIDFTPEWSLNLYQNVFLGDAIPV---GGSNEFAPVQNLTGSSVG  322

Qy   215  QTVILTWQAP----------ASDKRITVLNESFDTQLPNGWTMIDADGHNWLSTINVY   265
          |:||||:||             ::  ::||||:||||||||||||||||||||||||
Db   323  QKVFLKWDAPNGTPNPNPNPNPGTTLSESFE-NGIPASWKTIDADGDGHGW-KPGNAP  380

Qy   266  NTATHTGDGAMFSKSWTASGGAKIDLSPDNYLIVTPKVTVPENGKLSYWVSSQ-VPWTNEH  324
              :  :.::||:   ::||||||||||||||||||: :||
Db   381  GIAGYNSNGCVYSESF-GLGGIGV-LTPDNYLITPALDLPNGGKLTFWVCAQDANYASEH   438

Qy   325  YGVFLSTTGNEAANFTIKLLEETLGS----------DKPAPMNLV-  359
          .|||||||||||||||||||||||:
Db   439  YAVYASSTGNDASNFTNALLEETITAKGVRSPKAIRGRIQGTWRQKTVDLPAGTKYVAFR  498

Qy   360  -----------------KSEGVKL---------PAPYQERTID-  376
                            |:||:||           :|:
Db   499  HFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSG  558

Qy   377  ---LSAYAGQQV--  385
             |:|:
Db   559  QLDWLTAHGGSNVVSSFSWNGMALNPDNYLLSKDVTGATKVKYYAVNDGFPGDHYAVMI  618

Qy   386  -----------YLAFR  390
                       :||:
Db   619  SKTGTNAGDFTVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFR  678

Qy   391  HFNSTGIFRLYLDDV--AVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGGYNY  448
          |:||:||||         ||  |: ::|||:|||:|||||:|:||||: ::||||||||:
Db   679  HYNCSDLNYILLDDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEY  738

Qy   449  CVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAV--GQKVTLKWDAP  496
          ||||||||||||||||:||:||||||||||||||||    |:||||||||
Db   739  CVEVKYTAGVSPKKCVNVTV-NSTQFNPVQNLTAEQAPNSMDAILKWNAP  787
```

RESULT 13
Q9KIB3
ID   Q9KIB3   PRELIMINARY;   PRT;   312 AA.
AC   Q9KIB3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical outer membrane protein PG27.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C.;
RT   "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF237555; AAF81413.1; -.
SQ   SEQUENCE 312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;

Query Match   14.4%;   Score 379;   DB 2;   Length 312;
Best Local Similarity   37.2%;   Pred. No. 8.6e-16;
Matches 94;   Conservative 44;   Mismatches 85;   Indels 30;   Gaps 9;

```
Db   478  ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD  533
             ||| : :  :||||||||||||||  :  :             ||:   ::
Qy   295  NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---  350
          |||||||:||:||||:: : :|   ::||||   | | |||:  :   || :  || ::
Db   534  NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV  593

Qy   351  ----------------------DKPAPMNLV------KSEGVKL---------  366
                                  |:|:|||
Db   594  RSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET  653

Qy   367  ------PAPYQERTID-----------------LSAYAGQQV--------  385
               :|:
Db   654  FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSTNVVASFSWNGMALNPDNY  713

Qy   386  -----------------------------YLAFR  390
                                         :||:
Db   714  LISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFG   773

Qy   386  -----------------------------------YLAFR  390
                                               :||:
Db   774  LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT  417

Qy   418  YTVYRDNVVIAQNLAATTFNQENVAPGGYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV  477
          ||||||:|  :  ::|||||:  :|||||||:|:|||||||||||:||:||:||||||:|
Db   834  YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNPV  892

Qy   478  QNLTGSAV--GQKVTLKWDAP  496
          |||| :||    |:||||:||
Db   893  QNLTAEQAPNSMDAILKWNAP  913
```

RESULT 12
P72196
ID   P72196   PRELIMINARY;   PRT;   1097 AA.
AC   P72196;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   TonB-linked adhesin precursor.
GN   TLA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97386416; PubMed=9244265;
RA   Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA   Curtis M.A.;
RT   "The tla gene of Porphyromonas gingivalis W50: a homologue of the
RT   arginine-specific protease precursor (PrpRI) which shares sequence
RT   similarity to TonB-linked receptors.";
RL   J. Bacteriol. 179:4778-4788(1997).
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR   EMBL; Y07618; CAA68897.1; -.
DR   GO; GO:0019867; C:outer membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Membrane; Outer membrane; Receptor; Signal; TonB box.
FT   SIGNAL   1   53   POTENTIAL.
SQ   SEQUENCE 1097 AA; 118731 MW; 73BBA337B421F8B9 CRC64;

```
Qy  198  GGNEFAPVQNLQWS------------------------VSGGTVTLTWQAPASDK---------RTY-  231
         |:|:                                    :||:
Db  219  GQN---VGRLTWNYPEDYQPEGKGNEELQLSGYNIYANGTLLAQIKDVSILEVVDSTYS           274

Qy  232  -------------------------------------VLNESFDTQTLPNGWTMIDADGDGH          256
                                               :||:|||||
Db  275  LRDNPLQVEYCVTAVYDESIESSTVCGTLHYATDAILYENFENGPVPNGWLVIDADGDGF          334

Qy  257  NWLSTINVYNT-ATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGK-LSYWV          314
         ||||:||:|: :||||||||||||||||||||:|||| |||||:|||:   ||:|
Db  335  SWGHYLNAYDAFPGHNGGHCSLSASYVPGIG--PVTPDNYLITPKV--EGAKRVKYWV          388

Qy  315  SSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETLGSDK PAPMNLVKSEGVKLPAPYQER          373
         ||: ||||||:||||||||||||:::||||::||
Db  389  STQDANWAAEHYAVWMASTTGTAVGDFVI-LFEETM-TAKPT--------GAWYER          433

Qy  374  TIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTVYRDNVV--IAQNL          431
         |||:|||::|||||:|||||||||||||||||:::|||:
Db  434  TINLP--BGTK-YIAWRHYNCTDIYFLKLDDIIVFGTPASEPEPVT---DFVVSLIENNK          487

Qy  432  AATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKD---VTVBGSNEFA-----HVQNLT          481
         : :|                                :|:|:|:|:
Db  488  GRLKWN----------------------YPNGYEPDKTDDKDPLQLAGVNIYANGSLLVHIQDPT  530
```

```
RESULT 15
Q9XBU9
ID  Q9XBU9;         PRELIMINARY;         PRT;         293 AA.
AC  Q9XBU9;
DT  01-NOV-1999  (TrEMBLrel. 12, Created)
DT  01-NOV-1999  (TrEMBLrel. 12, Last sequence update)
DT  01-OCT-2002  (TrEMBLrel. 22, Last annotation update)
DE  Immunoreactive 32 kDa antigen PG25.
OS  Porphyromonas gingivalis (Bacteroides gingivalis).
OC  Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC  Porphyromonadaceae; Porphyromonas.
OX  NCBI_TaxID=837;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=W50;
RA  Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA  Hocking D., Webb E.;
RT  "P. gingivalis polypeptides and nucleic acids.";
RL  Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AF155351; AAD38410.1; -.
SQ  SEQUENCE  293 AA;  32272 MW;  CC03EAC241F7F6F1 CRC64;

Query Match                 8.1%;
Best Local Similarity       38.8%;     Score 213.5;  DB 2;  Length 293;
Matches    52;  Conservative    19;  Mismatches    42;  Indels    21;  Gaps    4;
```

```
Qy  362  EGVKLPAPYQER-TIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420
         ::||                                      :|:|
Db  39   ESLTFVAPVEETDAIEAEVEAIQEI-------------------VATEEIA---EQAVRSYTYTV  81

Qy  421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVBGSNEFAHVQNL  480
         |:||||                                 |||:||
Db  82   YRDGVKIASGLTEPTFLDEDVPAGEHTYCVEVQYQGGVSDKVCVDVEVK--DFKPVTNL  138

Qy  481  TGSAVGQKVTLKWD  494
         ||:|:
Db  139  TGTASNDEVSLDWD  152
```

```
Search completed: May 18, 2004, 11:46:34
Job time : 41.3342 secs
```

```
Qy  233  LNESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLS  292
         |:|||:: :||||||||| |||||||:||::|                    :||||
Db  4    LSESFES-GIPAIWKTIDADGDGYNWMHLTN-----FTGQSGLCVSSASYIGGVGA-LT  55

Qy  293  PDNYLVIPKVTVPENG--KLSYWVSSQ-VPWTNEHYGVFLSTGNEAANFTIKLLEETLG  349
         |||| :: |||                :||||||:||||||||||||||||||||:||
Db  56   PDNYLITPELKLPTDALVEIIYWVCTQDLTAPSEHYAVYSSSTGNNAADFVNLLYEETLT  115

Qy  350  SDKPAPMNLV--KSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVA  406
         :          |:                  |||:|||||||||||||||||||||||
Db  116  AKRIQSPELIRGNRTQGVW----YQRKVV----LPNDTKYVAFRHFNSTDNFWLNLDEVS  167

Qy  407  V-------SGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGV  458
                 :|:|:||||:|||:||            |: ||
Db  168  ILYTPLPRRAPCPHPGGYTYSVFRDGQKIASGLSALAYIDTDVPYGTQDYCVQVNVLQGD  227

Qy  459  SPKVCKDVTVEGS  471
         ||||:||
Db  228  SYKVCKNIVVANS  240
```

```
RESULT 14
Q9F4J0
ID  Q9F4J0;         PRELIMINARY;         PRT;         925 AA.
AC  Q9F4J0;
DT  01-MAR-2001  (TrEMBLrel. 16, Created)
DT  01-MAR-2001  (TrEMBLrel. 16, Last sequence update)
DT  01-MAR-2003  (TrEMBLrel. 23, Last annotation update)
DE  Putative outer membrane protein PG57.
OS  Porphyromonas gingivalis (Bacteroides gingivalis).
OC  Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC  Porphyromonadaceae; Porphyromonas.
OX  NCBI_TaxID=837;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=W50;
RA  Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA  Hocking D., Webb E.;
RT  "P. gingivalis polypeptides and nucleic acids.";
RL  Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=W50;
RA  Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA  Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT  "Identification of vaccine candidates from genomic analysis of
RT  Porphyromonas gingivalis.";
RL  Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AY007522; AAG24228.1; -.
DR  InterPro; IPR003961; FN_III.
DR  SMART; SM00060; FN3; 1.
SQ  SEQUENCE  925 AA;  103632 MW;  5FF2198D6914DAE8 CRC64;

Query Match                12.8%;
Best Local Similarity       25.4%;     Score 337;  DB 2;  Length 925;
Matches   152;  Conservative    62;  Mismatches   191;  Indels   194;  Gaps    31;
```

```
Qy  4    LNSLFSLAVLLSLLCW-----------GQTAAAQQGGPKTAPSVTHQAVQKGIRTSKVKD  51
         |: :                       :: ||   ::::
Db  5    MKSIVFRAFLITILLSWAAITNPTAQEISGMNASCLAAPAQPDTILYESFENG--------  56

Qy  52   LRDPIPAGMARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGT-----IPA  106
         ::||               |:::|         |||
Db  57   ---PVPNGWLEIDADA------DGA-----TWGSPSGSF--SVP----YGHNGLCTYSHIRS  98

Qy  107  GLYDPPEYKVPVNADA----------SFSPTNFVL---DGTASADIPAGTYDYVIIN  150
         |::|:  :                  |:  :|
Db  99   GISTAGNYLITPNIEGAKRVKYWVCNQYSTNPEHYAVMVSTTGTAIEDFVLLFDDSITGK  158

Qy  151  PNPGI----IYIVGEGVS-----KGNDYVVEAGKTYHFTVQR--QGPGDAASVVVTGE  197
         | :|     :||                   |:|:|:   |: :: ::|:|
Db  159  PTPLVWRRRIVDLPEGTKYIAWRHYKVTDSHTEFLKLDDVVYYRSIEGPEPATDFTVINI  218
```

```
              GenCore version 5.1.6
         Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:          May 18, 2004, 11:38:40 ; Search time 15.3555 Seconds
                                           (without alignments)
                                         1670.936 Million cell updates/sec

Title:           US-08-570-311-2
Perfect score:   2641
Sequence:        1 MRKLNSLFSLAVLLSLLCWG.........QNLTGSAVGQKVTLKWDAPN 497

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:   389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued Patents AA:*
                 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2641	100.0	497	2	US-08-570-311-2	Sequence 2, Appli
2	2641	100.0	497	2	US-08-353-485-2	Sequence 2, Appli
3	2641	100.0	2628	2	US-08-570-311-14	Sequence 14, Appli
4	951.5	36.0	1706	4	US-09-066-330-10	Sequence 10, Appl
5	944.5	35.8	1477	4	US-09-482-500A-1	Sequence 1, Appli
6	944.5	35.8	1704	3	US-08-336-308A-10	Sequence 10, Appli
7	944.5	35.8	1704	3	US-08-822-324-6	Sequence 6, Appli
8	944.5	35.8	1704	3	US-09-490-931-10	Sequence 10, Appl
9	943.5	35.7	1687	2	US-08-570-311-29	Sequence 29, Appl
10	816	30.9	1358	2	US-08-570-311-27	Sequence 27, Appl
11	812.5	30.8	1732	2	US-08-570-311-10	Sequence 10, Appl
12	812.5	30.8	1732	2	US-08-353-485-10	Sequence 10, Appl
13	806.5	30.5	1087	2	US-08-570-311-8	Sequence 8, Appli
14	806.5	30.5	1087	2	US-08-353-485-8	Sequence 8, Appli
15	806.5	30.5	1732	4	US-09-066-330-11	Sequence 11, Appli
16	546.5	20.7	450	2	US-08-570-311-16	Sequence 16, Appli
17	546.5	20.7	456	2	US-08-570-311-18	Sequence 18, Appli
18	546.5	20.7	456	2	US-08-570-311-20	Sequence 20, Appl
19	458.5	17.4	439	2	US-08-570-311-22	Sequence 22, Appli
20	405	15.3	942	1	US-08-141-324-14	Sequence 14, Appli
21	405	15.3	942	1	US-08-541-902-14	Sequence 14, Appli
22	175	6.6	49	3	US-08-822-324-18	Sequence 18, Appl
23	148	5.6	42	3	US-09-066-330-3	Sequence 3, Appli
24	140	5.3	2736	4	US-09-252-991A-30227	Sequence 30227, A
25	135.5	5.1	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
26	132	5.0	509	3	US-08-822-324-8	Sequence 8, Appli
27	127	4.8	46	3	US-08-822-324-9	Sequence 9, Appli
28	127	4.8	811	1	US-08-480-604A-7	Sequence 7, Appli
29	127	4.8	811	2	US-08-405-496A-7	Sequence 7, Appli
30	127	4.8	811	3	US-08-915-136-7	Sequence 7, Appli
31	127	4.8	811	4	US-08-957-310-7	Sequence 7, Appli
32	127	4.8	811	4	US-10-011-366-7	Sequence 7, Appli
33	127	4.8	811	4	US-09-084-517-7	Sequence 7, Appli
34	127	4.8	812	1	US-08-480-604A-29	Sequence 29, Appl
35	127	4.8	812	3	US-08-915-136-29	Sequence 29, Appl
36	127	4.8	812	3	US-09-084-517-29	Sequence 29, Appli
37	127	4.8	2710	1	US-08-480-604A-6	Sequence 6, Appli
38	127	4.8	2710	3	US-08-405-496A-6	Sequence 6, Appli
39	127	4.8	2710	3	US-08-915-136-6	Sequence 6, Appli
40	127	4.8	2710	4	US-08-957-310-6	Sequence 6, Appli
41	127	4.8	2710	4	US-10-011-366-6	Sequence 6, Appli
42	127	4.8	2710	4	US-09-084-517-6	Sequence 6, Appli
43	126	4.8	1752	4	US-09-865-621A-2	Sequence 2, Appli
44	123	4.7	951	2	US-09-924-097A-15	Sequence 15, Appl
45	121	4.6	25	2	US-08-902-516-46	Sequence 46, Appl

ALIGNMENTS

```
RESULT 1
US-08-570-311-2
; Sequence 2, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;    APPLICANT: Progulske-Fox, Ann
;    APPLICANT: Tumwasorn, Somying
;    APPLICANT: Lepine, Guylaine
;    APPLICANT: Han, Naiming
;    APPLICANT: Lantz, Marilyn
;    APPLICANT: Patti, Joseph
;    TITLE OF INVENTION:  Cloned Porphyromonas gingivalis Genes
;    TITLE OF INVENTION:  and Probes for the Detection of Periodontal Disease
;    NUMBER OF SEQUENCES: 29
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Ted W. Whitlock
;       STREET: 2421 N.W. 41st Street, Suite A-1
;       CITY: Gainesville
;       STATE: FL
;       COUNTRY: USA
;       ZIP: 32606
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/570,311
;       FILING DATE:
;       CLASSIFICATION: 424
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/353,485
;       FILING DATE: 09-DEC-1994
;       CLASSIFICATION: 424
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/647,119
;       FILING DATE: 25-JAN-1991
;       CLASSIFICATION: 424
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/241,640
;       FILING DATE: 08-SEP-1988
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Whitlock, Ted W.
;       REGISTRATION NUMBER: 36,965
;       REFERENCE/DOCKET NUMBER: UF15.C3
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (904) 375-8100
;       TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 2:
;    SEQUENCE CHARACTERISTICS:
```

```
;   LENGTH: 497 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-2

Query Match           100.0%;  Score 2641;  DB 2;  Length 497;
Best Local Similarity 100.0%;  Pred. No. 1.4e-205;
Matches 497;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM  60

Qy  61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120

Qy 121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180

Qy 181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240

Qy 241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300

Qy 301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK  360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK  360

Qy 361  SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361  SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420

Qy 421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480

Qy 481  TGSAVGQKVTLKWDAPN  497
        |||||||||||||||||
Db 481  TGSAVGQKVTLKWDAPN  497


RESULT 2
US-08-353-485-2
;  Sequence 2, Application US/08353485
;  Patent No. 5830710
GENERAL INFORMATION:
    APPLICANT: Progulske-Fox, Ann
    APPLICANT: Tumwasorn, Somying
    APPLICANT: Lepine, Guylaine
    APPLICANT: Han, Naiming
    APPLICANT: Lantz, Marilyn
    APPLICANT: Patti, Joseph
    TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
    TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
    NUMBER OF SEQUENCES: 12
    CORRESPONDENCE ADDRESS:
       ADDRESSEE: Ted W. Whitlock
       STREET: 2421 N.W. 41st Street, Suite A-1
       CITY: Gainesville
       STATE: FL
       COUNTRY: USA
       ZIP: 32606
    COMPUTER READABLE FORM:
       MEDIUM TYPE: Floppy disk
       COMPUTER: IBM PC compatible
       OPERATING SYSTEM: PC-DOS/MS-DOS
       SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
;   CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/353,485
;      FILING DATE: 09-DEC-1994
;      CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 07/647,119
;      FILING DATE: 25-JAN-1991
;      CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 07/241,640
;      FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;      NAME: Whitlock, Ted W.
;      REGISTRATION NUMBER: 36,965
;      REFERENCE/DOCKET NUMBER: UF15.C2
;   TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (904) 375-8100
;      TELEFAX: (904) 372-5800
;   INFORMATION FOR SEQ ID NO: 2:
;      SEQUENCE CHARACTERISTICS:
;         LENGTH: 497 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;         MOLECULE TYPE: Protein
US-08-353-485-2

Query Match           100.0%;  Score 2641;  DB 2;  Length 497;
Best Local Similarity 100.0%;  Pred. No. 1.4e-205;
Matches 497;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
US-08-353-485-2

Qy   1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM  60

Qy  61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120

Qy 121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180

Qy 181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVLNESFDTQ  240

Qy 241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300

Qy 301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK  360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK  360

Qy 361  SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361  SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420

Qy 421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480

Qy 481  TGSAVGQKVTLKWDAPN  497
        |||||||||||||||||
Db 481  TGSAVGQKVTLKWDAPN  497


RESULT 3
US-08-570-311-14
;  Sequence 14, Application US/08570311
;  Patent No. 5824791
;  GENERAL INFORMATION:
;      APPLICANT: Progulske-Fox, Ann
```

```
  ;   APPLICANT:  Tumwasorn, Somying
  ;   APPLICANT:  Lepine, Guylaine
  ;   APPLICANT:  Han, Naiming
  ;   APPLICANT:  Lantz, Marilyn
  ;   APPLICANT:  Patti, Joseph
  ;   TITLE OF INVENTION:  Cloned Porphyromonas gingivalis Genes
  ;   TITLE OF INVENTION:  and Probes for the Detection of Periodontal Disease
  ;   NUMBER OF SEQUENCES:  29
  ;   CORRESPONDENCE ADDRESS:
  ;     ADDRESSEE:  Ted W. Whitlock
  ;     STREET:  2421 N.W. 41st Street, Suite A-1
  ;     CITY:  Gainesville
  ;     STATE:  FL
  ;     COUNTRY:  USA
  ;     ZIP:  32606
  ;   COMPUTER READABLE FORM:
  ;     MEDIUM TYPE:  Floppy disk
  ;     COMPUTER:  IBM PC compatible
  ;     OPERATING SYSTEM:  PC-DOS/MS-DOS
  ;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
  ;   CURRENT APPLICATION DATA:
  ;     APPLICATION NUMBER:  US/08/570,311
  ;     FILING DATE:
  ;     CLASSIFICATION:  424
  ;   PRIOR APPLICATION DATA:
  ;     APPLICATION NUMBER:  US 08/353,485
  ;     FILING DATE:  09-DEC-1994
  ;     CLASSIFICATION:  424
  ;   PRIOR APPLICATION DATA:
  ;     APPLICATION NUMBER:  US 07/647,119
  ;     FILING DATE:  25-JAN-1991
  ;     CLASSIFICATION:  424
  ;   PRIOR APPLICATION DATA:
  ;     APPLICATION NUMBER:  US 07/241,640
  ;     FILING DATE:  08-SEP-1988
  ;   ATTORNEY/AGENT INFORMATION:
  ;     NAME:  Whitlock, Ted W.
  ;     REGISTRATION NUMBER:  36,965
  ;     REFERENCE/DOCKET NUMBER:  UF15.C3
  ;   TELECOMMUNICATION INFORMATION:
  ;     TELEPHONE:  (904) 375-8100
  ;     TELEFAX:  (904) 372-5800
  ;   INFORMATION FOR SEQ ID NO: 14:
  ;     SEQUENCE CHARACTERISTICS:
  ;       LENGTH:  2628 amino acids
  ;       TYPE:  amino acid
  ;       TOPOLOGY:  linear
  ;     MOLECULE TYPE:  protein
  ; US-08-570-311-14
```

```
Query Match             100.0%;    Score 2641;    DB 2;   Length 2628;
Best Local Similarity   100.0%;    Pred. No. 1.7e-204;
Matches  497;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy     1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  MRKLNSLFSLAVLLSLLCWGQTAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRDPIPAGM   60

Qy    61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  ARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDPFEYKVPVNA  120

Qy   121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT  180

Qy   181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVLTWQAPASDKRTYVLNESFDTQ  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  VQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVLTWQAPASDKRTYVLNESFDTQ  240

Qy   241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  TLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTP  300
```

```
Qy   301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  KVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVK  360

Qy   361  SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV  420

Qy   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480

Qy   481  TGSAVGQKVTLKWDAPN  497
          |||||||||||||||||
Db   481  TGSAVGQKVTLKWDAPN  497
```

```
RESULT 4
US-09-066-330-10
  ;   Sequence 10, Application US/09066330A
  ;   Patent No. 651666
  ;   GENERAL INFORMATION:
  ;   APPLICANT:  Reynolds, Eric C.
  ;   APPLICANT:  Bhogal, Peter S.
  ;   APPLICANT:  Slakeski, Nada
  ;   TITLE OF INVENTION:  DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
  ;   FILE REFERENCE:  Reynolds
  ;   CURRENT APPLICATION NUMBER:  US/09/066,330A
  ;   CURRENT FILING DATE:  1998-09-15
  ;   EARLIER APPLICATION NUMBER:  PN 6275
  ;   EARLIER FILING DATE:  1995-10-30
  ;   EARLIER APPLICATION NUMBER:  PCT/AU96/00673
  ;   EARLIER FILING DATE:  1996-10-30
  ;   NUMBER OF SEQ ID NOS: 15
  ;   SOFTWARE:  PatentIn Ver. 2.0
  ;   SEQ ID NO 10
  ;     LENGTH: 1706
  ;     TYPE: PRT
  ;     ORGANISM: Porphyromonas gingivalis
  ; US-09-066-330-10
```

```
Query Match             36.0%;    Score 951.5;    DB 4;   Length 1706;
Best Local Similarity   31.9%;    Pred. No. 5.6e-68;
Matches  237;   Conservative  74;   Mismatches 156;   Indels 275;   Gaps  20;

Qy    22  TAAAQG-------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH   68
          |:       ||||                :|||        ||| ||
Db   671  TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH  730

Qy    69  DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP  126
          ||||            |     : ::|  |||:   :  ||| :   ||  |:  ::
Db   731  DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP  789

Qy   127  TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG  185
          || :||||:|  |||||||::| :   ||     ||||:  | |||||||||    ||
Db   790  TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG  849

Qy   186  PGDAASVV------------------------------------------------    193
          |||:::||
Db   850  SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS  909

Qy   194  -----VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS--------------DKRTYVLNE  235
               |  || ||||:: :||||  :|||:|:|:|:                 |||||:||
Db   910  PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTLSE  969

Qy   236  SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN  295
          ||:  |:|||||:||||||||||:|||||||||||||||||||||:: |::  :| ||
Db   970  SFB-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1025

Qy   296  YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETL---GSD  351
          ||::|   :     :  |  :::  |    |  :  :  :|::|||||:|   :     ||
Db  1026  YLITPALDIPNGGKLTFWVCAQDANYASSHYAVVASSTGNDASNFTNALLEETITAKGVR 1085
```

```
Qy  352 KPAPM-------------------------------------------------------  356
Db 1086 SPEAMRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1145
Qy  357 -----NLVKS-----------------------------------                361
Db 1146 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSFSWNGMALNPDNYL  1205
Qy  362 -----------------------------------------------             361
Db 1206 ISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1265
Qy  362 ----EGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDV--AVSGEGSSND  415
Db 1266 STEADGAKPQSWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD  1322
Qy  416 YTYTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA  475
Db 1323 YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN 1381
Qy  476 HVQNLTGSAVGQKVTLKWDAPN 497
Db 1382 PVKNLKAQPDGDGDVVLKWEAPS 1403
```

RESULT 5
US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Imamura, Takahisa
; APPLICANT: Potempa, Jan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
; FILE REFERENCE: 235.00160101
; CURRENT APPLICATION NUMBER: US/09/482,500A
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/115,869
; PRIOR FILING DATE: 1999-01-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-482-500A-1

Query Match   35.8%; Score 944.5; DB 4; Length 1477;
Best Local Similarity   31.9%; Pred. No. 1.7e-67;
Matches   235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

```
Qy   22 TAAAGG------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH   68
Db  444 TATTGGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 503
Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
Db  504 DVWNDGSGYQILLDADHDQYKQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 562
Qy  127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
Db  563 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 622
Qy  186 PGDAASVV--                                                  193
Db  623 SGDGTELTISEGGSGDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 682
Qy  194 -----VTGEGGNEFAPVQNLQWSVSGQTVLTWQAPAS--------DKRTYVLNESF     237
Db  683 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSBSF   742
Qy  238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL 297
```

```
Db  743 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL  798
Qy  298 VTPKVTVPENGKLSYWYSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS------  350
Db  799 ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP  858
Qy  351 ---------------------DKPAPMNLV----------------KSEGVKL-------  366
Db  859 EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFBS  918
Qy  367 ----PAPYQERTID---------------------LSAYAGQQV---------------  385
Db  919 STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS  978
Qy  386 -----------------------------------------------            385
Db      -----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV                 420
Db  979 KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST 1038
Qy  386 -----------------------------------------------            386
Db 1039 EANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTV  1098
Qy  421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
Db 1099 YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1157
Qy  481 TGSAVGQKVTLKWDAPN 497
Db 1158 KAQPDGDGDVVLKWEAPS 1174
```

RESULT 6
US-08-336-308A-10
; Sequence 10, Application US/0833630BA
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
;   STREET: 5370 Manhattan Circle, Suite 201
;   CITY: Boulder
;   STATE: Colorado
;   COUNTRY: US
;   ZIP: 80303
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/336,308A
;   FILING DATE: 08-NOV-1994
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/119,361
;   FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/265,441
;   FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
;   NAME: Ferber, Donna M.
;   REGISTRATION NUMBER: 33,878
;   REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (303) 499-8080
;   TELEFAX: (303) 499-8089

```
;   INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1704 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-336-308A-10

Query Match    35.8%;   Score 944.5;  DB 3;  Length 1704;
Best Local Similarity  31.9%;  Pred. No. 2.1e-67;
Matches  235;  Conservative  75;  Mismatches 160;  Indels 267;  Gaps 20;

Qy    22 TAAAQG------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH   68
          ||     : |:    :  | :  ||   :  |||   : :||  | :|||||
Db   671 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH  730

Qy    69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP  126
          |||::| :||:| ||||::||  ||      :  ||| ||  | || ||  |||| |.||
Db   731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTIW-PNCSVPANLFAPFEYTVPENADPSCSP  789

Qy   127 TNFVLDGTASADIPAGTYDVVIINPNPGI-YIVGEGVSKGNDYVVEAGKTYHFTVQRQG  185
          || ::|||||:.|||||||. .|    :  || |:: | :|:|| ||||.|||.: :.|
Db   790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVPEAGKKYHFLMKKMG  849

Qy   186 PGDAASVV------                                               193
          .||:|:
Db   850 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATFEEDGVATGNHEYCVEVKYTAGVS  909

Qy   194 ------VTGEGNEFAPVQNLQWSVSGQTVLTWQAPAS------DKRTYVLNESF  237
                 ::|.||||||||| : .|| :||  |||             :|:||:||
Db   910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSESF  969

Qy   238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL  297
          : | | |  ||||||||||.|  : |     |   :    .  |.:| | ||.:|||||
Db   970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy   298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS------  350
          :||  .:| |||| ::| :   |||.||:|.: ||:|| : ||| :|||||:
Db  1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSP 1085

Qy   351 ------DKPAPMNLV------                                        366
                 ||:|:||:
Db  1086 EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLLDEVEIKANGKRADFTETFES 1145

Qy   367 --PAPYQERTID-----------LSAYAGQQV-------                      385
             |||  :|:           ||:||.||:
Db  1146 STHGEAPAEWTTIDADGDGGQWLCLSSGQLDWLIAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy   386 -------                                                      385
Db  1206 KDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST  1265

Qy   386 -------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV                 420
                  |||||| ||  |    |.|.    | |:|||:|||||
Db  1266 EANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTV 1325

Qy   421 YRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
          |||.. |:: |:.||| :: || | ::.|||||||||||||.| ||.|  :|||| ||||
Db  1326 YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNVPVKNL 1384

Qy   481 TGSAVGQKVTLKWDAPN                                            497
          :   | |.||:|||.||
Db  1385 KAQPDGGDVVLKWEAPS                                           1401
```

RESULT 7
US-08-822-324-6
; Sequence No. 6129917, Application US/08822324
; Patent No. 6129917
;   GENERAL INFORMATION:
;   APPLICANT: Potempa, Jan S.
;   APPLICANT: Travis, James
;   APPLICANT: Genco, Caroline A.

```
Matches   235;  Conservative   75;  Mismatches   160;  Indels   267;  Gaps   20;

Qy      22 TAAAQG--------GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH   68
              ||:|         :  |  :      |  :|       :|: :   ||.:   :|||
Db     671 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH  730

Qy      69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP  126
              |||||:||:|||| |||||  |||||  :.  :  :|: ::    |:|||: |:|||||
Db     731 DVWNDSSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP  789

Qy     127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKYHFTVQRQG  185
              ||:||||||:|||||||.|| :||||||  |  ||||   :||:|||||:|:|| |:|
Db     790 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKXYHFLMKKMG  849

Qy     186 PGDAASVV-                                                    193
              ||||:||
Db     850 SGDGTELTISEGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS  909

Qy     194 -----VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTYVLNESF  237
                   :|:| | | : : |  :|:|:  : || |||||        :  :||:|||
Db     910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSESF  969

Qy     238 DTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL  297
              : |    | | :|| |||.|  |  |:  | :|  ||||  :|  |    || :|||||
Db     970 E-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy     298 VTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS-----  350
              :||| ||||| |||  |:   |||| :|| ||:|| ||:|:| |:|||::  |
Db    1026 ITPALDLPNGGKLTFWVCAQDANYASHYAVYASSTGNDASNFTNALLEETTTAKGVRSP 1085

Qy     351 -------DKPAPMNLV-----------------------------KSEGVKL-------  366
                     :|||| :|:                              ||| |
Db    1086 EAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFES 1145

Qy     367 ---PAPYQERTID-----------------LSAYAGQQV-                    385
                 |||||||||                  ||:||| |
Db    1146 STHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLIS 1205

Qy     386 -------------------------------                             386
Db    1206 KDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST 1265

Qy     386 -------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV      420
                                 ||:||||||||||||||||  |||||||||||||||
Db    1266 EANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTV 1325

Qy     421 YRDNVVIAQNLAATTNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNL  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db    1326 YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVKNL 1384

Qy     481 TGSAVGQKVTLKWDAPN  497
              |||||||||||||:||:
Db    1385 KAQPDGGDVVLKWEAPS 1401
```

RESULT 8
US-09-490-931-10
; Sequence 10, Application US/09490931
; Patent No. 6274718
; GENERAL INFORMATION:
;   APPLICANT: Travis, James
;   APPLICANT: Potempa, Jan S.
;   APPLICANT: Barr, Philip J.
;   APPLICANT: Pavloff, Nadine
;   TITLE OF INVENTION: Porphyromonas gingivalis
;   TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
;   STREET: 5370 Manhattan Circle, Suite 201
;   CITY: Boulder
;   STATE: Colorado
;   COUNTRY: US
;   ZIP: 80303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/490,931
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/336,308
;   FILING DATE:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/265,441
;   FILING DATE: 24-JUN-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ferber, Donna M.
;   REGISTRATION NUMBER: 33,878
;   REFERENCE/DOCKET NUMBER: 21-93C
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (303) 499-8080
;   TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1704 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-490-931-10

Query Match   35.8%; Score 944.5; DB 3; Length 1704;
Best Local Similarity   31.9%; Pred. No. 2.1e-67;

RESULT 9
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tunwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville

```
Qy    367 -----------PAPYQERTID-------------LSAYAGQQV--------- 385
                     |:|: :||
Db   1127 ESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL 1186

Qy    386 -----------------------------------ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1246
                                          ||||||||
Db   1187 ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 1246

Qy    386 ------------------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTY 418
                                          :|||| :|||::|||| ||||
Db   1247 STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNVILLDDIQFTMGGSPTPDYTY 1306

Qy    419 TVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ 478
                |||||| :||||| |:||.| |:|
Db   1307 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTI-NPTQFNPVK 1365

Qy    479 NLTGSAVGQKVTLKWDAPN 497
                ||:|||| |||:|:||
Db   1366 NLKAQPDGGDVVLKWEAPS 1384
```

RESULT 10
US-08-570-311-27

; Sequence 27, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1358 amino acids

; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-29

Query Match   35.7%; Score 943.5; DB 2; Length 1687;
Best Local Similarity   31.8%; Pred. No. 2.5e-67;
Matches 235; Conservative 75; Mismatches 160; Indels 269; Gaps 20;

```
Qy     22 TAAAQG-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPIPAGMARIILEAH 68
                      ::|  :|                  :|::||||
Db    652 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 711

Qy     69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
              ||||:||  ||||| |  :|||:   |:::.|  |||||::|  ::|:||
Db    712 DVWNDGSGYQILLDADHDQYGQVIPSDTHILW-PNCSVPANLFAPFEYTVPENADPSCSP 770

Qy    127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQRQG 185
              ||::|| ||:| :|||||    :|:  :|||  : ||||::| ||||||  :::
Db    771 TNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG 830

Qy    186 PGDAASVV---------- 193
              || ||:
Db    831 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS 890

Qy    194 ------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------------DKRTYVLNE 235
                    ||||||||  ::|||||||||| ::||       ||:||||:
Db    891 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSE 950

Qy    236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
              :||  | |||  |:|||||||   ::::   :  ||:|  |:|: .:.|||||||| ||
Db    951 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1006

Qy    296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
              ||:|  :|:|  ||:||:|| |   ::|:|  :|||  ||     |::|:::|:
Db   1007 YLITPALDLFNGGKLTFWVCAQDANVASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1066

Qy    351 ----------DKPAPMNLV-------------------------KSEGVKL------- 366
                       |||||| |                        :| ||:|
Db   1067 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1126
```

```
;       TYPE: amino acid
;       TOPOLOGY: linear
;       MOLECULE TYPE: protein
US-08-570-311-27

Query Match           30.9%;  Score 816; DB 2; Length 1358;
Best Local Similarity 29.6%;  Pred. No. 3.7e-57;
Matches 219; Conservative 75; Mismatches 171; Indels 274; Gaps 21;

Qy   22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARIILEA 67
            |||    :.:  | |      |||.| |||          |||:::|
Db  328 TATTQGGQKVTLKWDAPS----AKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAA 382

Qy   68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
           :|:|.| |||:.| ||||:.:||:|||: |  |||||    ||:|:.|||||:|.|
Db  383 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 441

Qy  127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
           .:  ||.|: ||.||.| :| |::||    ||  ||:    :::|||: || ||||:||:|
Db  442 QNIIVTGQGEVVIPGGVVIDYCITNPEPASGKMWIAGDGDNQPARYDDFTFEAGKKYTFTM 501

Qy  182 QRQGPGDAASVV---------------------------- 193
            :
Db  502 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 561

Qy  194 ----------VTGEGGNEFAPVQNLQWSVSGGTVTLTWQAPAS--------DKRTYVLNE 235
                     ||.|..||||||||.| :|.:|.||| ||||            ::||||:|
Db  562 AGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTLSE 621

Qy  236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
           ||:     |.| ||||||||| |  :  .| ||  || :.:| |.    |||.| ||||
Db  622 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 677

Qy  296 YLIVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---- 350
           ||:  |     | ||| :|| :  .  ::|||.|:| ::|| |:||| .||||: :
Db  678 YLITPALDLANGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 737

Qy  351 --------DKPAPMNLV-------------------------KSEGVKL------- 366
                   :| | |.|                        ||.|.|
Db  738 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 797

Qy  367 ------PAPYQERTID----------------------LSAYAGQQV--------- 385
               | |.: |||                    |:| :||||
Db  798 ESSTHGEAPAEWTTIDADGDGDQDWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYL 857

Qy  386 ----------------------------------------- 385
Db  858 ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 917

Qy  386 ---------YIAFRHFNSTGIPIRFRLYLDDV--AVSGEGSSNDYTY 418
                   ||||||.|:| ||| .|||:|:|   |.||..|.||||
Db  918 STEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYILLDDIQFTMGGSPTPTDYTY 977

Qy  419 TVYRDNVVIAQNLAATTFNQENVAPGQVNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQ 478
           |||||:  |: |: || ||:|:| || :|||||||||||||||||| |:|. :||:| |
Db  978 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNPVK 1036

Qy  479 NLTGSAVGQKVTLKWDAPN 497
           || .   |..||||| ||
Db 1037 NLKAQPGGGDVVLKWEAPS 1055
```

```
;       TITLE OF INVENTION:  Cloned Porphyromonas gingivalis Genes
;       TITLE OF INVENTION:  and Probes for the Detection of Periodontal Disease
;       NUMBER OF SEQUENCES: 29
;       CORRESPONDENCE ADDRESS:
;          ADDRESSEE:  Ted W. Whitlock
;          STREET:  2421 N.W. 41st Street, Suite A-1
;          CITY:  Gainesville
;          STATE:  FL
;          COUNTRY:  USA
;          ZIP:  32606
;       COMPUTER READABLE FORM:
;          MEDIUM TYPE:  Floppy disk
;          COMPUTER:  IBM PC compatible
;          OPERATING SYSTEM:  PC-DOS/MS-DOS
;          SOFTWARE:  PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;          APPLICATION NUMBER:  US/08/570,311
;          FILING DATE:
;          CLASSIFICATION:  424
;       PRIOR APPLICATION DATA:
;          APPLICATION NUMBER:  US 08/353,485
;          FILING DATE:  09-DEC-1994
;          CLASSIFICATION:  424
;       PRIOR APPLICATION DATA:
;          APPLICATION NUMBER:  US 07/647,119
;          FILING DATE:  25-JAN-1991
;          CLASSIFICATION:  424
;       PRIOR APPLICATION DATA:
;          APPLICATION NUMBER:  US 07/241,640
;          FILING DATE:  08-SEP-1988
;       ATTORNEY/AGENT INFORMATION:
;          NAME:  Whitlock, Ted W.
;          REGISTRATION NUMBER:  36,965
;          REFERENCE/DOCKET NUMBER:  UF15.C3
;       TELECOMMUNICATION INFORMATION:
;          TELEPHONE:  (904) 375-8100
;          TELEFAX:  (904) 372-5800
INFORMATION FOR SEQ ID NO: 10:
;       SEQUENCE CHARACTERISTICS:
;          LENGTH: 1732 amino acids
;          TYPE: amino acid
;          TOPOLOGY: linear
;       MOLECULE TYPE: protein
US-08-570-311-10

Query Match           30.8%;  Score 812.5; DB 2; Length 1732;
Best Local Similarity 30.0%;  Pred. No. 1e-56;
Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;

Qy   22 TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARIILEA 67
            |||     :.:  | |      |||.| |||         |||:::|
Db  693 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 747

Qy   22 TAAAQGGPKT----APSVTHQAVQKGIRTSKVKDLRD--------PIPAGMARIILEA 67
            |||     :.:  | |      |||.| |||         |||:::|
Db  693 TATTQGGQKVTLKWEAPS-----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy   68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
           :|:|.| |||:.| ||||:.:||:|||: |  |||||    ||:|:.|||||:|.|
Db  748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVTT 806

Qy  127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
           .:  ||.|: ||.||.| :| |::||    ||  ||:    :::|||: || ||||:||:|
Db  807 QNIIVTGQGEVVIPGGVVDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM 866

Qy  182 QRQGPGDAASVV---------------------------- 193
            :
Db  867 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy  194 ----------VTGEGGNEFAPVQNLQWSVSGGTVTLTWQAP-------ASDKRTYVLN 234
                     ||.|..||||||||.| :|.:|.||| ||||
Db  927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS 986

Qy  235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
           ||:     |.| ||||||||| |  :  .| ||  || :.:| |.    |||.| ||||
Db  987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042
```

```
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1732 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-353-485-10

  Query Match             30.8%;  Score 812.5;  DB 2;  Length 1732;
  Best Local Similarity   30.0%;  Pred. No. 1e-56;
  Matches  222;  Conservative   70;  Mismatches  172;  Indels  277;  Gaps   22;

Qy     22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD----PIPAGMARILEA  67
             |||||     |||    :  ::||     ::||    :
Db    693 TATTGQKVTLKWFAPS---AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy     68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
           ||| ::|| ||  | ||||| || ||||      |||||| ||  | || || ||| ||
Db    748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVTT 806

Qy    127 TNFVLDGTASADIPAGTYDVIINPNP--GIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
           :|  | :|   ||| | ||  | |||    | :||        :|     || ||
Db    807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDFFTFEAGKKYTFTM 866

Qy    182 QRQGPGDAASVV-- 193
           :|    ||  ::
Db    867 RRAGMGDGTDMEVEDDSPASYTYVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy    194 -----VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP---------ASDKRTYVLN 234
                |||  ||||||||||  :  | | ||  :|
Db    927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS 986

Qy    235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
           ||| :  | :| | |||||||| | | :   |         |   :| :    || |||
Db    987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042

Qy    295 NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS--- 350
           |||:||  ::|:||||: ||  |   : :|||| | ||:|| | ||| ||||||   |
Db   1043 NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102

Qy    351 -----------DKPAPMNLV---------KSEGVKL----- 366
                      |:  || |          ||  ||
Db   1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy    367 -----PAPYQERTID----------LSAYAGQQV----- 385
                |||  | ||           ||  ||  |
Db   1163 FESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy    386 ----- 385
Db   1223 LISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFG 1282

Qy    386 -----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
                ||||||| :| |  || |||    ||| | :|:||
Db   1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNVILLDDIQFTMGGSPTPTDYT 1342

Qy    418 YTVYRDNVVIAQNLAATFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
           ||||||  :|:  | :| ||:  | | | | |||||||||||| | ||||  ||  |:|
Db   1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVDVTV-NSTQFNPV 1401

Qy    478 QNLTGSAV--GQKVTLKWDAP 496
           ||||   |   | | ||| ||
Db   1402 QNLTAEQAPNSMDAILKWNAP 1422
```

RESULT 12
US-08-353-485-10
; Sequence 10, Application US/08353485
; Patent No. 5830710
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800

RESULT 13
US-08-570-311-8
; Sequence 8, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine

```
Db  371  GW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYLIITPALDLANGGKLITFWVCA  427
              :  ::  ::|||  : :||||| ::| |: |||| ::||  ::
Qy  317  Q-VPWTNEHYGVFLSTGNEAANFTKLLEETLGS------DKP  353
              :  :||:|| |:|||:| ||||| :
Db  428  QDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLP  487

Qy  354  APMNLV-----------KSEGVKL----------PAPYQERTID-----  376
              |  ::     :  ::
Db  488  AGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDG  547

Qy  377  ------LSAYAGQQV-----  385
              ||  : ::|:
Db  548  QDWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGF  607

Qy  386  ------  385

Db  608  PGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLP  667
Qy  386  ----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTTVYRDNVVIAQNLAATTFNQE  439
              :||||||:|:| | ||:|:|||  :  ::||: ||||||| ::||| || ::
Db  668  AGTKYVAFRHYNCSDLDYILLDDIQTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEED  727

Qy  440  NVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQKVTLKWDAPN  497
              || ||:|||||||||||||||||||| :| :||||||:||| :|||||||||| ||
Db  728  GVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNPVKNLKAQPDGGDVVLKWEAPS  784
```

RESULT 14
US-08-353-485-8

US-08-570-311-8

SEQUENCE CHARACTERISTICS:
LENGTH: 1087 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

```
Query Match         30.5%;   Score 806.5;  DB 2;  Length 1087;
Best Local Similarity  29.4%;  Pred. No. 1.6e-56;
Matches 211;  Conservative 76;  Mismatches 166;  Indels 265;  Gaps 19;

Qy   39  AVQKGIRTSKVKDLRD---------PIPAGMARIILEAHDVWEDGTGYQMLWDADHNQY   88
              :  ::  :  |  :|  |||  :|: ::|||::||:| :||||| ||||||:
Db   73  SAKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTF  132
Qy   89  GASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASPSPTNFVLDGTASADIPAGTYDYV  147
              |:  |   :|| ||||||| || ||||| |||| :  :  ||||  :|| |||| |||
Db  133  GSVIPATGPLF-TGIASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVYDYC  191
Qy  148  IINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTVQRQGPGDAASVV---------  193
              ||| |   ::||  :|: | |||| | :||||:|:||  : : |
Db  192  ITNPEPASGKMWLAGDGDNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASY  251
Qy  194  ----------------------VTGEGGNEFAP  204
                                    :||| ||||||
Db  252  TYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAP  311
Qy  205  VQNLQWSVSGQTVTLTWQAPAS--------DKRTYVLNESFDTQTLPNGWTMIDADGDGH  256
              ||||  |  ||:|||:| ||       |: :| :||:|||: :  ||||:|||||||
Db  312  VQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTTLSESFE-NGIPASWKTIDADGDGH   370
Qy  257  NWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYWVSS  316
```

```
;  SEQUENCE CHARACTERISTICS:
;    LENGTH: 1087 amino acids
;    TYPE: amino acid
;    TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-353-485-8

Query Match     30.5%;   Score 806.5;  DB 2;  Length 1087;
Best Local Similarity   29.4%;  Pred. No. 1.6e-56;
Matches 211;  Conservative 76;  Mismatches 166;  Indels 265;  Gaps 19;

Qy   39 AVQKGIRTSKVKDLRD---------PIPAGMARIILEAHDVWEDGTGYQMLWDADHNQY 88
            : :  ::|||:|              ||||||||||||||||||
Db   73 SAKKAEASREVKRIGDGLFVTIEPANDVRANEAKVVLAADNVWGNTGYQFLLDADHNTF 132

Qy   89 GASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP-TNFVLDGTASADIPAGTYDYV 147
         ||:|  ||| |:||||:: | ::|||:|||| | ||  :||:|:| ||: ||||:|||:
Db  133 GSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTTQNIIVTGQGEVVIPGGVYDYC 191

Qy  148 IINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTVQRQGPGDAASVV-- 193
         |.||.|  | :| :| :    ::  |  :|||| | |:  | |:||  :    
Db  192 ITNPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASY 251

Qy  194 --------VTGEGQNEFAP 204
                  | || ||||||
Db  252 TYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAP 311

Qy  205 VQNLQWSVSGQTVTLTWQAPAS------DKRTYVLNESFDTQTLPNGWTMIDADGDGH 256
         ||||  | |||:|||:|:||         ::  |.|||: ||| ||:|||||||||:|
Db  312 VQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTLSESFE-NGIPASWKTIDADGDGH 370

Qy  257 NWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYWVSS 316
            |::|  | | : ||  :| |: | |: |:  |:||||:|||:| ||: ||||::|::
Db  371 GW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDNYLITPALDLANGGKLTFWVCA 427

Qy  317 Q-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---------DKP 353
         : : : :|:|| ::::||:|| ||| :||||| :          | 
Db  428 QDANYASEHYAVVASSTGNDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLP 487

Qy  354 APMNLV----------KSEGVKL----------- 376
         |   :            | |:|            
Db  488 AGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDG 547

Qy  377 -----------LSAYAGQQV------------- 385
                     :| |||:|               
Db  548 QDWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLLISKDVTGATKVKYYYAVNDGF 607

Qy  386 ------- 385
                
Db  608 PGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLP 667

Qy  386 ----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 439
             || ||:| : | :  |||||    | |:| |: |:
Db  668 AGTKYVAFRHYNCSDLDYLLLDDIQFTMGGSPTPTDYTVYRDGTKIKEGLTETTFEED 727

Qy  440 NVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQKVTLKWDAPN 497
            | | :|||||||||||||||| ||.| ||||||: ||||:||: |:|||||:|| 
Db  728 GVATGNHEYCVEVKYTAGVSPKVCVNVTI-NPTQFNPVKNLKAQPDGGDVVLKWEAPS 784
```

```
;  EARLIER FILING DATE: 1995-10-30
;  EARLIER APPLICATION NUMBER: PCT/AU96/00673
;  EARLIER FILING DATE: 1996-10-30
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 11
;    LENGTH: 1732
;    TYPE: PRT
;    ORGANISM: Porphyromonas gingivalis
US-09-066-330-11

Query Match     30.5%;   Score 806.5;  DB 4;  Length 1732;
Best Local Similarity   29.7%;  Pred. No. 3.2e-56;
Matches 220;  Conservative 72;  Mismatches 172;  Indels 277;  Gaps 22;

Qy   22 TAAAQGGPKT---APSVTHQAVQKGIRTSKVKDLRD----------PIPAGMARIILEA 67
          | |:| | ||         : :  ::|||:|              |||||||||||
Db  693 TATTQGQKVTLKWEAPS----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

Qy   68 HDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
          :||| ||||:|| |||||:| ||:| |||  ||:: ||| |  :|||:|:||||| |:
Db  748 DNVWGNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 806

Qy  127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTV 181
          :|:|:| ||: ||||:|||:|.||.|  | :| :| :    ::  |  :|||| | |:
Db  807 QNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWIAGDGNQPARYDDFTFEAGKKYTFTM 866

Qy  182 QRQGPGDAASVV-- 193
          | |:||  :   
Db  867 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT 926

Qy  194 ---------VTGEGQNEFAPVQNLQWSVSGQTVTLTWQAP------ASDKRTYVLN 234
                   | || |||||||||  | |||:|||:|:||        ::  | |.|
Db  927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPNPNPGTTLS 986

Qy  235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD 294
         |||: ||| ||:||||||||:|   |::|  | | : ||  :| |: | |: |:  |:|
Db  987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042

Qy  295 NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS--- 350
         |||:|||:| ||: ||||::|::: : : :|:|| ::::||:|| ||| :||||| :
Db 1043 NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNALLEETITAKGV 1102

Qy  351 --------------DKPAPMNLV----------- 366
                       | |   :            
Db 1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy  367 --------PAPYQERTID--------- 385
                  |   |:||             
Db 1163 FESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY 1222

Qy  386 ------- 385
                
Db 1223 LISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFG 1282

Qy  386 --------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
                       || ||:| : | :  |||||    | |:| |: |:
Db 1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNVILLDDIQFTMGGSPTPTDYT 1342

Qy  418 YTVVRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV 477
         ||||||  | :|:|: ||| : |  ||  |||||||||||||||| ||.| ||||||: 
Db 1343 YTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFNPV 1401

Qy  478 QNLTGSAV-GQKVTLKWDAP 496
         ||||:|||  |:|||||:|
Db 1402 QNLTAEQAPNSMDAILKWNAP 1422

Search completed: May 18, 2004, 11:49:12
Job time : 21.3555 secs
```

```
RESULT 15
US-09-066-330-11
;  Sequence 11, Application US/09066330A
;  Patent No. 6511666
;  GENERAL INFORMATION:
;  APPLICANT: Reynolds, Eric C.
;  APPLICANT: Bhogal, Peter S.
;  APPLICANT: Slakeski, Nada
;  TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;  FILE REFERENCE: Reynolds
;  CURRENT APPLICATION NUMBER: US/09/066,330A
;  CURRENT FILING DATE: 1998-09-15
;  EARLIER APPLICATION NUMBER: PN 6275
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:42:50 ; Search time 38.497 Seconds
                            (without alignments)
                            3592.387 Million cell updates/sec

Title: US-08-570-311-2
Perfect score: 2641
Sequence: 1 MRKLNSLFSLAVLLSLLCWG.........QNLTGSAVGQKVTLKWDAPN 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications AA:*
    1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
    2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
    3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
    4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
    5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
    6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
    7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
    8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
    9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
   10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
   11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
   12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
   13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
   14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
   15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
   16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
   17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
   18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	951.5	36.0	1706	14	US-10-229-066-10	Sequence 10, Appl
2	806.5	30.5	1732	14	US-10-229-066-11	Sequence 11, Appl
3	791.5	30.0	419	15	US-10-174-695-3	Sequence 3, Appli
4	662.5	25.1	231	15	US-10-174-695-5	Sequence 5, Appli
5	457.5	17.3	196	15	US-10-174-695-6	Sequence 6, Appli
6	188.5	7.1	29	15	US-10-174-695-4	Sequence 4, Appli
7	150	5.7	29	15	US-10-387-977-27	Sequence 27, Appl
8	150	5.7	1483	12	US-10-282-122A-51483	Sequence 51483, A
9	148	5.6	42	14	US-10-229-066-3	Sequence 3, Appli
10	145	5.5	509	15	US-10-387-977-101	Sequence 101, App
11	144	5.5	31	15	US-10-387-977-6	Sequence 6, Appli
12	143	5.4	1946	12	US-10-282-122A-62947	Sequence 62947, A
13	141	5.3	31	15	US-10-387-977-9	Sequence 9, Appli
14	139	5.3	26	15	US-10-387-977-15	Sequence 15, Appl
15	138.5	5.2	2468	12	US-10-282-122A-66335	Sequence 66335, A

16	138.5	5.2	2468	14	US-10-246-330-4	Sequence 4, Appli
17	137	5.2	25	15	US-10-387-977-82	Sequence 82, Appl
18	137	5.2	579	15	US-10-369-493-9075	Sequence 9075, Ap
19	134	5.1	2435	12	US-10-282-122A-47453	Sequence 47453, A
20	132	5.0	698	14	US-10-156-761-11953	Sequence 11953, A
21	129	4.9	973	14	US-10-156-761-9394	Sequence 9394, Ap
22	128	4.8	26	15	US-10-387-977-14	Sequence 14, Appl
23	127	4.8	811	12	US-10-272-898-7	Sequence 7, Appli
24	127	4.8	811	14	US-10-011-366-7	Sequence 7, Appli
25	127	4.8	811	14	US-10-354-774-7	Sequence 7, Appli
26	127	4.8	811	15	US-10-271-012-7	Sequence 7, Appli
27	127	4.8	2710	14	US-10-272-898-6	Sequence 6, Appli
28	127	4.8	2710	14	US-10-011-366-6	Sequence 6, Appli
29	127	4.8	2710	15	US-10-354-774-6	Sequence 6, Appli
30	127	4.8	2710	15	US-10-271-012-6	Sequence 6, Appli
31	126	4.8	1752	14	US-10-387-388-2	Sequence 2, Appli
32	125.5	4.8	577	15	US-10-369-493-21739	Sequence 21739, A
33	124	4.7	2013	12	US-10-282-122A-60608	Sequence 60608, A
34	124	4.7	2358	12	US-10-282-122A-45763	Sequence 45763, A
35	123.5	4.7	1204	12	US-10-282-122A-49627	Sequence 49627, A
36	123	4.7	866	14	US-10-222-038-2	Sequence 2, Appli
37	123	4.7	951	9	US-09-924-097-15	Sequence 15, Appl
38	122	4.6	806	15	US-10-369-493-3678	Sequence 3678, Ap
39	121	4.6	25	9	US-09-847-185-46	Sequence 46, Appl
40	121	4.6	25	10	US-09-930-915A-18	Sequence 18, Appl
41	121	4.6	25	14	US-10-224-286-46	Sequence 46, Appl
42	121	4.6	25	14	US-10-082-014-40	Sequence 40, Appl
43	121	4.6	25	14	US-10-372-076-41	Sequence 41, Appl
44	121	4.6	25	15	US-10-387-977-76	Sequence 76, Appl
45	121	4.6	25	15	US-10-387-977-79	Sequence 79, Appl

ALIGNMENTS

RESULT 1
US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US2003015637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match   36.0%; Score 951.5; DB 14; Length 1706;
Best Local Similarity   31.9%; Pred. No. 3.5e-75;
Matches 237; Conservative   74; Mismatches 156; Indels 275; Gaps   20;

Qy   22 TAAAQG------GPKTAPSVTHQAVOK-GIR---TSKVKDLRDPIPAGMARIILEAH 68
              || : ||:|| :| | :   :|||:||::|:||||||||
Db   671 TATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 730

Qy   69 DVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFBYKVPVNADASFSP 126
              |||||:| ::|:||||:||:|||||| :|: ||| ||:||:|:||:|||||||
Db   731 DVWNDGSGYQILLDADHDQYGQVIPSDTHTLW-PNCSVPANLFAPFBYTVPENADPSCSP 789

Matches 220; Conservative 72; Mismatches 172; Indels 277; Gaps 22;

```
Qy    22 TAAAQQGPKT---APSVTHQAVQKGIRTSKVKDLRD---------PIPAGMARIILEA   67
            :||:|
Db   693 TATTQQGQKVTLKWEAPS----AKKAEGSREVKRIGDGLFVTIEPANDVRANEAKVVLAA  747

Qy    68 HDVWEDGTGYQMLWDADHNQYGASIPESFWFANGTIPAGLYDP-FEYKVPVNADASFSP  126
          :|| || ||||| || ||| | ||:|| ||   |   | |  ||  ::|||||  ::
Db   748 DNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT  806

Qy   127 TNFVLDGTASADIPAGTYDYVIINPNP---GIYIVGEG---VSKGNDYVVEAGKTYHFTV  181
          :|:::|||||  |||||| | :|:|| :  |:| ||     ::|||  | |||| |:||
Db   807 QNLIVTGQGEVVIPGGVYDYCIITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTM  866

Qy   182 QROGPGDAASVV-----------------------------------            193
          :
Db   867 RRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVEVKYT  926

Qy   194 --------------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAP------ASDKRTYVLN  234
                        |||:||||||||||| : ||:||:|||||
Db   927 AGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKDAPNGTPNPNPNPNPGTTLS     986

Qy   235 ESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPD  294
          |||: | | |||:|||||||||| |    |   | | ||    |  | :|| |  ||:||
Db   987 ESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPD 1042

Qy   295 NYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETLGS---  350
          |||: ||  |||||||::||  | |  | |||| :  |||||:||| :||||||
Db  1043 NYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102

Qy   351 --------------DKPAPMNLV------------------KSEGVKL-             366
                        | ||
Db  1103 RSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTET 1162

Qy   367 -------PAPYQBRTID---------LSAYAGQQV-                          385
                 | | ::|||         |||:|| |
Db  1163 FESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGSNVVSSFSWNGMALNPDNY  1222

Qy   386 -----------------------------------                          385
Db  1223 LISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFG  1282

Qy   386 ----------------------YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT      417
                                ::||||::|   |::|| |   |  |  |  :||
Db  1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT 1342

Qy   418 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHV  477
          |||||| | :|:||: ||| :||| ||:::|||||||||||||:|  :|||:||||||
Db  1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFNPV 1401

Qy   478 QNLTGSAV-GQKVTLKWDAP                                         496
          |||| :|| ||:|||||:|||
Db  1402 QNLTAEQAPNSMDAILKWNAP                                       1422
```

```
Qy   127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQROG  185
          :|:::||||| :||||||| |   |  :: |:| || |:|:|| | |||| |||  :::
Db   790 TNMIMDGTASVNIPAGTYDFALAAPQANAKIWIAGQGPTKEDDYVPEAGKKYHFLMKKMG  849

Qy   186 PGDAASVV-----                                                193
          ||
Db   850 SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVS  909

Qy   194 -------VTGEGGNEFAPVQNLQWSVSGQTVTLTWQAPAS---------DKRTYVLNE    235
                 |||:||||||||||| : ||:||:|||||
Db   910 PKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPGTENPNPNPNPNPGTTTLSE    969

Qy   236 SFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDN 295
          ||: | | |||:|||||||||| |    |   | | ||    |  | :|| |  ||:|||
Db   970 SFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GLGGIGV-LTPDN 1025

Qy   296 YLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLLEETL---GSD  351
          ||: ||  |||||||::||  | |  | |||| :  |||||:||| :||||||
Db  1026 YLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1085

Qy   352 KPAPM-----------                                             356
          | ||
Db  1086 SPEAMRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF 1145

Qy   357 ---------NLVKS                                               361
                   ::|
Db  1146 ESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVSSFSWNGMALNPDNYL  1205

Qy   362 ---------------                                              361
Db  1206 ISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL  1265

Qy   362 ---EGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDV--AVSGEGSSND  415
             : || | |::|||:||:|| ||::||||::|   |::|| |   |  |  |  :||
Db  1266 STEADGAKPQSVWIERTVDLP--AGTK-YVAFRHYNCSDLNYILLDDIQFTMGGSPTPTD 1322

Qy   416 YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFA   475
          |||| | | :|:||: ||| :||| ||:::|||||||||||||:|  :|||:||||||
Db  1323 YTVYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTV-NSTQFN 1381

Qy   476 HVQNLTGSAVGQKVTLKWDAPN                                       497
           :|||  | |||:|||||:||
Db  1382 PVKNLKAQPDGGDVVLKWEAPS                                      1403
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

Query Match       30.0%;  Score 791.5; DB 15; Length 419;
Best Local Similarity   39.9%;  Pred. No. 8e-62;
Matches  174;  Conservative  61;  Mismatches 100;  Indels 101;  Gaps  14;

Qy     58 AGMARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEE--SFWFANGTIPAGLYDPFEYK 115
          :|||||||||||||||||||| :||||||| ||:||| | |:||||||||||| |||
Db      1 SGQAEIVLEAHDVWNDGSGYQILLDADHDQYGVIPSDTHTLW-PNCSVPANLFAPFEYT  59

Qy    116 VPVNADASFSPTNFVLDGTASADIPAGTYDVIINPNGI-IYIVGEGVSKGNDYVVEAG 174
          ||:||| || |:|:||  |:|:|||:  :|| |||||||| ||||
Db     60 VPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAG 119

Qy    175 KTYHFTVQRGPGDAASVV------------------------ 193
          |:|||| :|||||| :||
Db    120 KKYHFLMKKMGSGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEY 179

Qy    194 ------------VTGEGGNEFAPVQNLQWSVSGQTVTLITWQAPAS-------------- 226
                       |||||||||||||||||||||||||||||:
Db    180 CVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSAVGQKVTLKWDAPNGTPNPNPNP 239

Qy    227 --DKRTYVLNESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTAS 284
            :|||||||||||  :|:||  :|:|:
Db    240 NPNPGTTLSESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GL 296

Qy    285 GGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKL 343
          |||||||||||:|||| |:|:||  :|||::|:|||||||||||||||||||||||||||
Db    297 GGIGV-LTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 355

Qy    344 LEETLGSDKPAPMNLVKSEGVKLP------APYQERTIDLSAYAGQQVYLAFRHFNST 395
          ||||                         :|:|||||||||||||||||||||||||||
Db    356 LEET------------ITAKGVRSPEAIRGRIQGTWRQKTVDLP--AGTK-YVAFRHFQST 401

Qy    396 GIFRLYLDDVAVSGEG 411
          :|:||||||||||||
Db    402 DMFYIDLDEVEIKANG 417

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US2003023022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-5

Query Match       25.1%;  Score 662.5; DB 15; Length 419;
Best Local Similarity   36.0%;  Pred. No. 2.7e-50;
Matches  157;  Conservative  60;  Mismatches 118;  Indels 101;  Gaps 15;
```

```
Qy     58 AGMARIILEAHDVWEDGTGYQMLWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKV 116
          | :|||:|| : ||| :|:||| ||| |||||| |||||||||||:|||:||
Db      1 ANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLI  59

Qy    117 PVNADASFSPTNFVLDGTASADIPAGTYDVIINPNP--GIIYIVGEG---VSKGNDYVV 171
          | || |:|| |  :|||  ::|  :| : |:| :  :::
Db     60 PANADPVVTQNIIVTGQGEVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTF 119

Qy    172 EAGKTYHFTVQRGPGDAASVV 193
          |||:|||| ::|||||:||||
Db    120 EAGKKYTFTMRRAGMGDGTMEVEDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGN 179

Qy    194 --------------VTGEGGNEFAPVQNLQWSVSGQTVTLITWQAP-------------- 224
                       |||||||||||||||||||||||||||||
Db    180 HEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKWDAPNGTPNPNPN 239

Qy    225 ASDKRTYVLNESFDTQTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTAS 284
          :|| ||||||||||||||  :|:|| |:|:|
Db    240 PNPNPGTTLSESFE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSESF-GL 296

Qy    285 GGAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQ-VPWTNEHYGVFLSTTGNEAANFTIKL 343
          ||||||||||||:|||| |:|:||  :|||::|:|||||||||||||||||||||||||:
Db    297 GGIGV-LTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 355

Qy    344 LEETLGSDKPAPMNLVKSEGVKLP------APYQERTIDLSAYAGQQVYLAFRHFNST 395
          |||                          :|:|||||||||||:||||||||||||||
Db    356 LEET------------ITAKGVRSPKAIRGRIQGTWRQKTVDLP--AGTK-YVAFRHFQST 401

Qy    396 GIFRLYLDDVAVSGEG 411
          :|:||||||||||||
Db    402 DMFYIDLDEVEIKANG 417

RESULT 5
US-10-174-695-6
; Sequence 6, Application US/10174695
; Publication No. US2003023022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-6

Query Match       17.3%;  Score 457.5; DB 15; Length 231;
Best Local Similarity   27.8%;  Pred. No. 2.2e-32;
Matches  118;  Conservative  22;  Mismatches  63;  Indels 221;  Gaps   6;

Qy     80 LWDADHNQYGASIPEESFWFANGTIPAGLYDP-FEYKVPVNADASFSPTNFVLDGTASAD 138
          |:||||| : : |||| |||||||||||:|||
Db      2 LLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTQNIIVTGQGEVV  60

Qy    139 IPAGTYDVIINPNP--GIIYIVGEG---VSKGNDYVVEAGKTYHFTVQRGPGDAASVV 193
          |||| : |:| :  ::: |||:||||
Db     61 IPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDME 120

Qy    194 VTGEGGNEFAPVQNLQWSVSGQTVTLITWQAPASDKRTYVLNESFDTQTLPNGWTMIDADG 253
          ||||||||||||||||||||||||||||
Db    121 VE------------------------------------------------------- 122
```

```
Qy  254  DGHNWLSTINVNTATHTGDGAMFSKSWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYW  313

Db  123  ---------------------------------------------------------  122

Qy  314  VSSQVPWTNEHYGVFLSTTGNEAANFTIKLLEETLGSDKPAPMNLVKSEGVKLPAPYQER  373
                                                 |||
Db  123  ------------------------DDSPA----------------------------  127

Qy  374  TIDLSAYAGQQYYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTVYRDNVVIAQNLAA  433
                                                   ||||||||     :  |
Db  128  -------------------------------SYTYTVYRDGTKIKEGLTA           146

Qy  434  TTFNQENVAPGQYNYCVEVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQKVTLKW  493
          |||   :: ||  :|| :||||||||||||||||||||||||||||   |||||||:|||||||||
Db  147  TTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQKVTLKW  206

Qy  494  DAPN  497
          ||||
Db  207  DAPN  210
```

RESULT 6
US-10-174-695-4
; Sequence 4, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-4

Query Match   7.1%; Score 188.5; DB 15; Length 196;
Best Local Similarity 33.5%; Pred. No. 1.8e-08;
Matches 59; Conservative 8; Mismatches 50; Indels 59; Gaps 5;

```
Qy  194  VTGEGGNEFAPVQNLQWSVSGQTVLTWQAPAS------------DKRTYVLNESFDTQT  241
          || |||||||||||  :||  |||| ||  :|:|||  ||
Db    6  VTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPGTTLSESFE-NG        64

Qy  242  LPNGWTMIDADGDGHNWL--------------------------STINVYNTATHTGDG  274
          :|||| ||||||||  :
Db   65  IPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCVLDNSAKIDRNQEINVYNTAEY----  120

Qy  275  AMFSKSWTASGGAKIDLSP--DNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGV      327
          ||  :||        |  : :|           ||
Db  121  --------AKTNNAPIKVVGYADEKTGTAAYNMKLSERRAKAVAKMLEKYGV        164
```

RESULT 7
US-10-387-977-27
; Sequence 27, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-27

Query Match   5.7%; Score 150; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy  415  DYTYTVYRDNVVIAQNLAATTFNQENVAP  443
          |||||||||||||||||||||||||||||
Db    1  DYTYTVYRDNVVIAQNLAATTFNQENVAP  29
```

RESULT 8
US-10-282-122A-51483
; Sequence 51483, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERNCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 51483

```
; LENGTH: 1483
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51483

Query Match     5.7%;  Score 150;  DB 12;  Length 1483;
Best Local Similarity  20.2%;  Pred. No. 0.0011;
Matches 127;  Conservative  81;  Mismatches 213;  Indels 208;  Gaps 29;

Qy    27  GGPKTAP------SVTHQAVQKGIRTSKVKDLRDPIPAGMARIILEAHDVWEDCTGY-  77
Db   522  GTPVVTPSQINVEQGSATDQPVKIDLNGNTLKDVDQ----SGKTILVQGIDYTVTDTGIT  577

Qy    78  ------QMLWDADHNQYGAS--------IPEESFWPANGTIPAGLYDPFEYKVPV-  118
Db   578  LSQSYLAGLALGQYTILTLDFNGGGASQTITINVVKNETVKLSVGTVSGNPGD--TVKVPV  635

Qy   119  N-------ADASFSPTNF-VLDGTASADIPAGT--YDYVIINPNPGIIYIV-----  159
Db   636  TISQVSTPVGLICMDISYDASKFTVKDVLPNTDLVKDTDNYSFIVNTSTPGKISITFTDP  695

Qy   160  ------GEGVSKGNDYVVEAGKT--YHFTVQRQGPGDAASVVTGEGNEFAPVQNL  208
Db   696  TLANYPISVDGILAYLDFIINSNATRAGDSALTV------DPATLIVADEND------KDI  743

Qy   209  QWSVSGQTVTLTWQAPASDKRTYVLNESFDT--QTLP---NGWTMIDA-DGDGH  256
Db   744  KDAASNGKITVTGSAPV--VQSSVVNTSSVTYDQNAPQDQAVSITFNGNTVKDVKDASG-  800

Qy   257  NWLSTINVYNTATHTGDGAMFSKSWTASGGA-----KIDLSPDN----YLVTPKVTVPE  306
Db   801  ---NTLKAGSDYTATSDGITLSQSYLATLAAGTYTVTIDFSAGNAGTFTVVVKGKTVVGS  857

Qy   307  NGKLSYWVSS------QVPWT-----NEHYGVFLSTTGNEAANFTIK-------LLEET-  347
Db   858  ATTLAVGTVSGKAGDTVKPVPVTISKVTTPVGLICAEIDYDASKFTVKDVLPNTDLVKDTD  917

Qy   348  -----LGSDKPAPMNL---------VKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRH  391
Db   918  NYSFIVNTSTPGKISITFTDPTLANYPISADGI---LAYLDFIINSNATAGDSALTV--  971

Qy   392  FNSTGIFRLYLDDVAVSGEGSSNDYT------  417
Db   972  -NPSGFIIADENDKDIQDAASNGKIVTGSTPVAENSVVNTSSVTYDQNAPQDQAVSITL  1030

Qy   418  ------YTVYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKY-T  455
Db  1031  NGNTITDVKDASGNTLKAGSDYTVTSDGITLSQSYLAT------LAAGTYTYTVDFSAGN  1084

Qy   456  AGVSPKVCKDVTVEGSNEFAHVQNLTGSA  484
Db  1085  AGTFTVVVKAKTVVSSATTLAVGTVSGKA  1113

RESULT 9
US-10-229-066-3
; Sequence 3, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
;   APPLICANT:  Reynolds, Eric C.
;   APPLICANT:  Bhogal, Peter S.
;   APPLICANT:  Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE:  2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR APPLICATION FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR APPLICATION FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR APPLICATION FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15


; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-3

Query Match     5.6%;  Score 148;  DB 14;  Length 42;
Best Local Similarity  64.1%;  Pred. No. 7.3e-06;
Matches 25;  Conservative  7;  Mismatches  7;  Indels  0;  Gaps  0;

Qy    58  AGMARIILEAHDVWEDCTGYQMLWDADHNQYGASIPEES  96
Db     1  SGQAEIVLEAHDVWNDGSGYQILLDADHDQYGQVIPSDT  39

RESULT 10
US-10-387-977-101
; Sequence 101, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
;   APPLICANT:  Reynolds, Eric Charles
;   APPLICANT:  O'Brien-Simpson, Neil Martin
;   APPLICANT:  Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION:  DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION:  PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200030I
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE:  2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR APPLICATION FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR APPLICATION FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR APPLICATION FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-101

Query Match     5.5%;  Score 145;  DB 15;  Length 509;
Best Local Similarity  19.1%;  Pred. No. 0.00061;
Matches 102;  Conservative  65;  Mismatches 155;  Indels 212;  Gaps 22;

Qy   114  YKVPVN--ADASF------------SPTNFVLD---------GTASADIPAGTY----D  145
Db    10  YNTPVRMLVVAGAKFKEALKPWLTWKAQKGFYLDVHYTDEAEVGTTNASIKAFIHKKYND  69

Qy   146  YVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFTVQRQGPGDAASVVTGEGNEFAPV  205
Db    70  GLAASAAPVFLALVGD------------TDVISGEKGKKTKKV  100

Qy   206  QNLQWS-VSG----QTVTLTWQAPASDKRTYVLNE--SFDTQTLPN---GWTMIDADGD  254
Db   101  TDLYYSAVDGDYFPEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAGAD  160

Qy   255  GHNWLSTI----------------NVYN-----TATHTGDGAMFSKSWTASGG  286
Db   161  -YSWNSQVGQPTIKYGMQYYYNQEHGYTDVYNYLKAPVTGCYSHLNTGVSFA-NYTAHGS  218

Qy   287  A------------KIDLSPDNYLVTPK------VTVPENGKLSYWV  314
Db   219  ETAWADPLLTTSQLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRVKEKGAYAYIG  278

Qy   315  SSQVPWTNEHY---------GVFLSTTGNEAANFTIKLLEET----  347
Db   279  SSPNSYWGBDYYWSVGANAVFGVYQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH  338

Qy   348  ------------LGSDKPAPMN-LVKSEGVKLPA--PYQERTIDLSAYAGQ  383
```

```
Db   339  AGNIGNITHIGAHYYWEAYHVLGDGSVMPYRAMPKNTYTLPASLPQNQASYSIQASAGS  398
Qy   384  QVYLAPRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTVYRDNVVIAQNLAATTFNQENVAP 443
Db   399  Y--------------VAISKDG-------VLYGTGVANASGVATVSMTKQITEN       431
Qy   444  GQYNYCVBVKYTAGVSPKVCKDVTVEGSNEFAHVQNLTGSAVGQKVTLKWDAPN       497
Db   432  GNY----DVVITRSNYLPVIKQIQVGEPSPYQPVSNLTATTQGQKVTLKWEAPS       481
```

RESULT 11
US-10-387-977-6
; Sequence 6, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
APPLICANT: Reynolds, Eric Charles
APPLICANT: O'Brien-Simpson, Neil Martin
APPLICANT: Slakeski, Nada
TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
FILE REFERENCE: 5292820000301
CURRENT APPLICATION NUMBER: US/10/387,977
CURRENT FILING DATE: 2003-07-18
PRIOR APPLICATION NUMBER: US 09/423,056
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: PCT/AU98/00311
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: AU PO 6528
PRIOR FILING DATE: 1997-04-30
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 6
LENGTH: 31
TYPE: PRT
ORGANISM: Porphyromonas gingivalis
US-10-387-977-6

Query Match   5.5%; Score 144; DB 15; Length 31;
Best Local Similarity   96.6%; Pred. No. 1e-05;
Matches   28; Conservative   0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   469  EGSNEFAHVQNLTGSAVGQKVTLKWDAPN  497
          |||||||||||||||||||||||||||||
Db     1  EGSNEFAHVQNLTGSAVGQKVTLKWDAPN   29
```

RESULT 12
US-10-282-122A-62947
; Sequence 62947, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 62947
LENGTH: 1946
TYPE: PRT
ORGANISM: Moraxella catarrhalis
US-10-282-122A-62947

Query Match   5.4%; Score 143; DB 12; Length 1946;
Best Local Similarity   18.4%; Pred. No. 0.0071;
Matches 110; Conservative 63; Mismatches 214; Indels 210; Gaps 23;

```
Qy    22  TAAAQGGPKTAPSVTHQAVQKGIRTSKVKDLRD---PIPAGMARIILEAHDVWEDGTGY      77
Db   417  TATLLAGGLTVKDIHNKQIQVSADGIKFTDMNGGTSAPVAAGTTRIT------RDTIGF    469
Qy    78  QMLWDADHNQYGASIPESSFWFANGTIPAGLYDPFEYKVPVNADASFSPTNFVLDG-TAS   136
Db   470  A------KEGGTVDDNSPYLDNAKLKVG------NVEINKDTGINAGNQKITGLTDG     514
Qy   137  ADIPAGTYDVIINPNPGIIYIVGEGVSGKGNDYVVEAGKTYHFTVQRQGPGDAASVVVTG   196
Db   515  ADDADAVTIKQLKNAKPNL--TPGNGIEINN------TNSLLVDA                551
Qy   197  EGGNEFAPVQNLQWSVSGQTVTLTWQAPA------SDKRTY------VLNESFD        238
Db   552  ANGNVTPP---SYTIGVKTTLTLNNGPSDSKFAVKDSNTHNNSLVTAKDLAGYLNEVNR   607
Qy   239  T-QTLPNGWTMIDADGDGHNWLSTINVYNTATHTGDGAMFSKSWTASGGAKIDLSPDNYL   297
Db   608  TADSALQSFTVKGGDAASNNSITLLDKTSNSLTVAGDNGITVKIDTTQKKVTVGIDQAANGL 667
Qy   298  VTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTGNEAANFTIKLLEETL------      348
Db   668  TTFKLTVGSNSKTQLVIEQMV------IGNDTKGNI-IKGLSSTLTDITNTDNT        713
Qy   349  ------GSDKPA-----PMNLVKSEGVKLPAPYQ                            371
Db   714  HTTEQDNDAQGKESNAASIKDVLNAGFNLQGNGKPVDFVSTYDVNFANGDGTATVTY-    772
Qy   372  ERTIDLSAYAGQQYLAFRHFN----STGIFRLYLDDVAV-----SGEGSSNDYTYTVY    421
Db   773  ------NYNEGKPTSTVVYNVNVDDTTIHLANNNGNKKLSVKTTTLT               813
Qy   422  RDNVVIAQNLAATTF------NQENVAPGQYNYCVEVKYTAGVSP-----KVCKD-     465
Db   814  KTN--GANGNATKFSANNGDALVNAKGIADNLNILAEEIHTTKGTADTALQTFKVKKDG   870
Qy   466  ------VTVEGSNEFA-----HVQNLTGSAVGQKVTLKWD                      494
Db   871  ATNDDDTITVGKNAVDTLIFKGENGLTVATKKDGTVTFGINTQSGLKAGDNTTLNKD     927
```

RESULT 13
US-10-387-977-9

; Sequence 9, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928O000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-9

Query Match   5.3%;   Score 141;   DB 15;   Length 31;
Best Local Similarity   93.1%; Pred. No. 1.9e-05;
Matches 27; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   469 EGSNEFAHVQNLTGSAVGQKVTLKWDAPN 497
             ||||| |||||||||:||||||||||||||
Db   1 EGSNEFAPVQNLTGSSVGQKVTLKWDAPN 29

RESULT 14
US-10-387-977-15
; Sequence 15, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928O000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-15

Query Match   5.3%;   Score 139;   DB 15;   Length 26;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   202 FAPVQNLQWSVSGQTVTLTWQAPASD 227
             ||||||||||||||||||||||||||
Db   1 FAPVQNLQWSVSGQTVTLTWQAPASD 26

RESULT 15
US-10-282-122A-66335
; Sequence 66335, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66335
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66335

Query Match   5.2%;   Score 138.5;   DB 12;   Length 2468;
Best Local Similarity   23.7%; Pred. No. 0.026;
Matches 107; Conservative   47; Mismatches 166; Indels 131; Gaps   23;

Qy   103 TIPAGLYDPFEYKVPVNADASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEG 162
             ||:|| ||| :|| || :|
Db   319 TDPAG------NNSTPVTVEAP--------DTTA----PAPATD---------VQVAPDG 351

Qy   163 VS-KGNDYVVEAGKTYHFTVQRQGPGDAASVVVTGEGGNEFAPVQNLQWSVSGQTVTLTW 221
             || :||| | ||| :||| |:|:|:
Db   352 SSVTGN---AEPGAT--VGVDTDGDGQPDTTVVVGPGGSFEVPLN--PPLTNGETVTVIV 404

Qy   222 QAPASDKRTYVLNESFDTQTLPN--------------GWTMIDADGDGH---------- 256
             |:::|||| :|: |||
Db   405 TDPAGNSSTPVTAEAPDFFDAPQVNASNGSVLSGTAEAGVTIVITDGNGNPIGQTSADAN 464

Qy   257 -NWLST----------INVY------NTATH-TGDGAM------------FSKSWTASG 285
             :|: ||:
Db   465 GNWSFTPGSQLPDGTVNVVARDAAGNSSPATSITVDGVAPNAPVVEPSNGSELSGTAEP 524

Qy   286 GAKIDLSPDNYLVTPKVTVPENGKLSYWVSSQVPWTNEHYGVFLSTTGNEAANFTIKLLE 345
                               |:   ::| :| :
Db   525 GSSVTLTDGNGNPIGQTTADANGNWSFTPSTFLP-----DGTVVNVVARDAAGNSSPPAS 579

```
Qy    346 ETLGSDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDV 405
              |:::||   ||   :   |    :: :||  |:: ||   :|--------|
Db    580 VTVDAVAPATPTVDPSNGT--------TLSGTAEPGSSVTLTDGNGNPIG--------- 621

Qy    406 AVSGEGSSNDYTYT----VYRDNVVIAQNLAATTFNQENVAPGQYNYCVEVKYTAGVSPK 461
              |:::||| :|:|    :     ||  ||     ||    ||   :|  :|
Db    622 QVTADGSGN-WTFTPSTPLPNGTVV---NATATDPSGNASSPAS----VTVDAVAPATPV 673

Qy    462 V--CKDVTVEGSNEFAHVQNLT---GSAVGQ 487
             |:  |||||||       |:  ||: :||
Db    674 VNPSNGTTLSGTAEPGATVTLTDGNGNPIGQ 704


Search completed: May 18, 2004, 11:52:19
Job time : 40.497 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:        May 18, 2004, 11:33:39 ; Search time 46.4099 Seconds
                                        (without alignments)
                                        2739.638 Million cell updates/sec

Title:          US-08-570-311-16
Perfect score:  2443
Sequence:       1 PNPNPGTTLSESFENGIPA.........QNLTGSAVGQXVTLKWDAPN 450

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1586107 segs, 282547505 residues

Total number of hits satisfying chosen parameters:       1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      A_Geneseq_29Jan04:*
                1: geneseqp1980s:*
                2: geneseqp1990s:*
                3: geneseqp2000s:*
                4: geneseqp2001s:*
                5: geneseqp2002s:*
                6: geneseqp2003as:*
                7: geneseqp2003bs:*
                8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2443	100.0	450	2	AAR96021	Aar96021 P. gingiv
2	2443	100.0	450	2	AAW69489	Aaw69489 Haemagglu
3	2443	100.0	2628	2	AAR96030	Aar96030 P. gingiv
4	2443	100.0	2628	2	AAW69488	Aaw69488 Haemagglu
5	2436	99.7	456	2	AAR96023	Aar96023 P. gingiv
6	2436	99.7	456	2	AAR96022	Aar96022 P. gingiv
7	2436	99.7	456	2	AAW69491	Aaw69491 Haemagglu
8	2436	99.7	456	2	AAW69490	Aaw69490 Haemagglu
9	2321	95.0	439	2	AAR96024	Aar96024 P. gingiv
10	2321	95.0	439	2	AAW69492	Aaw69492 Haemagglu
11	2063	84.4	1687	2	AAR96033	Aar96033 P. gingiv
12	2063	84.4	1687	2	AAW69495	Aaw69495 Haemagglu
13	2063	84.4	1704	2	AAR70188	Aar70188 Arg-gingi
14	2063	84.4	1704	2	AAW34843	Aaw34843 Arg-gingi
15	2063	84.4	1704	3	AAY67396	Aay67396 Arg-gingi
16	2063	84.4	1704	4	AAU08938	Aau08938 P. gingiv
17	2051	84.0	1706	2	AAW24786	Aaw24786 PrtR anti
18	2036	83.3	1087	2	AAR96028	Aar96028 P. gingiv
19	2036	83.3	1087	2	AAW69486	Aaw69486 Haemagglu
20	2036	83.3	1358	2	AAR96032	Aar96032 P. gingiv
21	2036	83.3	1358	2	AAW69494	Aaw69494 Haemagglu
22	2011.5	82.3	1732	2	AAR96029	Aar96029 P. gingiv
23	2011.5	82.3	1732	2	AAW24787	Aaw24787 PrtK anti
24	2011.5	82.3	1732	2	AAW69487	Aaw69487 Haemagglu
25	727	29.8	135	6	ABP55081	Abp55081 Porphyrom
26	711	29.1	134	4	AAB49217	Aab49217 Peptide u
27	700	28.7	419	4	AAU03572	Aau03572 P. gingiv
28	683	28.0	921	2	AAY34522	Aay34522 Porphyrym
29	683	28.0	922	2	AAY34521	Aay34521 Porphorym
30	683	28.0	925	2	AAY34520	Aay34520 Porphorym
31	683	28.0	938	2	AAY34392	Aay34392 Porphorym
32	680.5	27.9	419	4	AAU03574	Aau03574 P. gingiv
33	565	23.1	377	2	AAY34359	Aay34359 Porphorym
34	561.5	23.0	312	2	AAY34484	Aay34484 Porphorym
35	546.5	22.4	497	2	AAR96025	Aar96025 P. gingiv
36	546.5	22.4	497	2	AAW69483	Aaw69483 Haemagglu
37	415	17.0	970	2	AAR72458	Aar72458 Porphyrom
38	385	15.8	231	4	AAU03575	Aau03575 P. gingiv
39	267	10.9	991	2	AAR77313	Aar77313 Porphyrom
40	236	9.7	49	2	AAW34805	Aaw34805 Arg-speci
41	233	9.5	293	2	AAY34483	Aay34483 Porphorym
42	233	9.5	299	2	AAY34358	Aay34358 Porphorym
43	206.5	8.5	196	4	AAU03573	Aau03573 P. gingiv
44	189	7.7	46	2	AAW34798	Aaw34798 Arg-speci
45	153	6.3	29	2	AAW83085	Aaw83085 Peptide f

```
                              ALIGNMENTS

RESULT 1
AAR96021
ID   AAR96021 standard; protein; 450 AA.
XX
AC   AAR96021;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep1 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep1.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;  95WO-US016108.
XX
PR   09-DEC-1994;  94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30645.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 103-104; 153pp; English.
XX
CC   HArep1 (AAR96021) is the product of the HArep1 repeat unit (AAT30645) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep1 and other hagA repeat
CC   unit products (see also AAR96022-24) can be obtd. from transformed host
CC   cells and used as vaccines to protect humans or animals against
CC   periodontal disease. Expression in Salmonella cells allows prodn. of live
CC   vaccine. HArep1-4 can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 450 AA;
```

XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 450 AA;

Query Match   100.0%; Score 2443; DB 2; Length 450;
Best Local Similarity   100.0%; Pred. No. 1.2e-184;
Matches 450; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60

Qy   61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120

Qy   121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180

Qy   181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240

Qy   241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300

Qy   301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360

Qy   361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420

Qy   421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
              ||||||||||||||||||||||||||||||
Db   421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450

RESULT 3
AAR96030
ID   AAR96030 standard; protein; 2628 AA.
XX
AC   AAR96030;
XX
DT   16-OCT-2003   (revised)
DT   04-SEP-1996   (first entry)
XX
DE   P. gingivalis haemagglutinin hagA.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
FH   Key   Location/Qualifiers
FT   Peptide   5..21
FT   /label= Sig_peptide
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995; 95WO-US016108.
XX
PR   09-DEC-1994; 94US-00353485.

KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58875.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 91-110; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 2628 AA;

Query Match   100.0%;   Score 2443; DB 2;   Length 2628;
Best Local Similarity   100.0%;   Pred. No. 1.4e-183;
Matches 450;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy       1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     500 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI 559

Qy      61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     560 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 619

Qy     121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     620 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 679

Qy     181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     680 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 739

Qy     241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     740 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 799

Qy     301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     800 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 859

Qy     361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     860 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 919

Qy     421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
           ||||||||||||||||||||||||||||||
Db     920 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 949
```

XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30654.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 6; Page 93-101; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagA (AAR96030) was identified as the
CC   product of the hagA gene (AAT30654) isolated as an EcoRV fragment of
CC   genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
CC   24), can be obtd. from transformed host cells and used as a vaccine to
CC   protect humans or animals against periodontal disease. Expression in
CC   Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
CC   also be used to detect the presence of anti-P. gingivalis antibodies and
CC   to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-
CC   2003 to standardise OS field)
XX
SQ   Sequence 2628 AA;

Query Match   100.0%;   Score 2443; DB 2;   Length 2628;
Best Local Similarity   100.0%;   Pred. No. 1.4e-183;
Matches 450;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy       1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     500 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI 559

Qy      61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     560 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 619

Qy     121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     620 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 679

Qy     181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     680 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 739

Qy     241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     740 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 799

Qy     301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     800 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 859

Qy     361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     860 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 919

Qy     421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
           ||||||||||||||||||||||||||||||
Db     920 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 949
```

RESULT 4
AAW69488
ID   AAW69488 standard; protein; 2628 AA.
XX
AC   AAW69488;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagA.
XX

```
Db  307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVARRHYNCSDLNYILLDDIQFT 366
         ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Qy  361 MGGSPPTPDTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  367 MGGSPPTPDTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 426

Qy  421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
         ||||||||||||||||||||||||||||||
Db  427 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

RESULT 5
AAR96023
ID   AAR96023 standard; protein; 456 AA.
XX
AC   AAR96023;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep3 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep3.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30647.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 110-112; 153pp; English.
XX
CC   HArep3 (AAR96023) is the product of the HArep3 repeat unit (AAT30647) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep3 and other hagA repeat
CC   unit products (see also AAR96021-22 and AAR96024) can be obtd. from
CC   transformed host cells and used as vaccines to protect humans or animals
CC   against periodontal disease. Expression in Salmonella cells allows prodn.
CC   of live vaccine. HArep1-4 can also be used to detect the presence of anti
CC   -P. gingivalis antibodies and to raise monoclonal antibodies for
CC   diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 456 AA;

Query Match   99.7%; Score 2436; DB 2; Length 456;
Best Local Similarity   99.8%; Pred. No. 4.2e-184;
Matches 449; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGWLCLSSGQLDWLTAHGGTNVVASFS 60
         ||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   7 PNPNPGTTLSESFENGIPASWKTIDADGDGWLCLSSGQLGWLTAHGGTNVVASFS 66

Qy   61 NFEGPQNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKIGTNAGDFTVFEETP 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67 NFEGPQNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKIGTNAGDFTVFEETP 126

Qy 121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 186

Qy 181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASFS 240
         |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db 187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTNVVASFS 246

Qy 241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKIGTNAGDFTVFEETP 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKIGTNAGDFTVFEETP 306

Qy 301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
```

RESULT 6
AAR96022
ID   AAR96022 standard; protein; 456 AA.
XX
AC   AAR96022;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep2 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep2.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A,  Tumwasorn S,  Lepine G,  Han N,  Lantz M,  Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30646.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 107-108; 153pp; English.
XX
CC   HArep2 (AAR96022) is the product of the HArep2 repeat unit (AAT30646) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep2 and other hagA repeat
CC   unit products (see also AAR96021 and AAR96023-24) can be obtd. from
CC   transformed host cells and used as vaccines to protect humans or animals
CC   against periodontal disease. Expression in Salmonella cells allows prodn.
CC   of live vaccine. HArep1-4 can also be used to detect the presence of anti
CC   -P. gingivalis antibodies and to raise monoclonal antibodies for
CC   diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 456 AA;

Query Match             99.7%;  Score 2436;  DB 2;  Length 456;
Best Local Similarity   99.8%;  Pred. No. 4.2e-184;
Matches  449;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 66

Qy   61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126

Qy  121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 186
             ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Qy      181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
             |||||||||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db      187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFS 246
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      247 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      367 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 426
             ||||||||||||||||||||||||||||||
Qy      421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
             ||||||||||||||||||||||||||||||
Db      427 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

```
RESULT 7
AAW69491
ID   AAW69491 standard; protein; 456 AA.
XX
AC   AAW69491;
XX
DT   22-DEC-1998  (first entry)
XX
DE   Haemagglutinin protein hagA, HArep3.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58878.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 133-138; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 456 AA;

Query Match          99.7%;   Score 2436;   DB 2;   Length 456;
Best Local Similarity 99.8%;   Pred. No. 4.2e-184;
Matches  449;   Conservative  0;   Mismatches  0;   Indels  1;   Gaps  0;
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        7 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 66
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       67 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 186
             ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Qy      181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
             |||||||||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db      187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFS 246
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      247 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      367 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 426
             ||||||||||||||||||||||||||||||
Qy      421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
             ||||||||||||||||||||||||||||||
Db      427 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

```
RESULT 8
AAW69490
ID   AAW69490 standard; protein; 456 AA.
XX
AC   AAW69490;
XX
DT   22-DEC-1998  (first entry)
XX
DE   Haemagglutinin protein hagA, HArep2.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58877.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 127-132; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
```

CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 456 AA;

Query Match   99.7%;   Score 2436; DB 2; Length 456;
Best Local Similarity   99.8%;   Pred. No. 4.2e-184;
Matches   449; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI   66

Qy   61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126

Qy   121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 186

Qy   181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 246

Qy   241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306

Qy   301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366

Qy   361 MGGSPTPTDYTYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 MGGSPTPTDYTYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 426

Qy   421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
            ||||||||||||||||||||||||||||||
Db   427 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456

RESULT 9
AAR96024
ID   AAR96024 standard; protein; 439 AA.
XX
AC   AAR96024;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep4 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep4.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.

DR   N-PSDB; AAT30648.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 114-115; 153pp; English.
XX
CC   HArep4 (AAR96024) is the product of the HArep4 repeat unit (AAT30648) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep4 and other hagA repeat
CC   unit products (see also AAR96021-23) can be obtd. from transformed host
CC   cells and used as vaccines to protect humans or animals against
CC   periodontal disease. Expression in Salmonella cells allows prodn. of live
CC   vaccine. HArep1-4 can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 439 AA;

Query Match   95.0%;   Score 2321; DB 2; Length 439;
Best Local Similarity   98.4%;   Pred. No. 5e-175;
Matches   426; Conservative   3; Mismatches   4; Indels   0; Gaps   0;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db   7 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICVSSASYI   66

Qy   61 NFEGPQNPDNYLVTPELS:PNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
            |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   67 NFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126

Qy   121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||::||||||||
Db   127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIK 186

Qy   181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 246

Qy   241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306

Qy   301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366

Qy   361 MGGSPTPTDYTYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 MGGSPTPTDYTYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNV 426

Qy   421 TVDPVQFNPVQNL 433
            |::|:||||||||
Db   427 TINPTQFNPVQNL 439

RESULT 10
AAW69492
ID   AAW69492 standard; protein; 439 AA.
XX
AC   AAW69492;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagA, HArep4.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.

XX
PF 11-DEC-1995;   95US-00570311.
XX
PR 08-SEP-1988;   88US-00241640.
PR 25-JAN-1991;   91US-00647119.
PR 09-DEC-1994;   94US-00353485.
XX
PA (UYFL) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR WPI; 1998-582627/49.
DR N-PSDB; AAV58879.
XX
PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT protease poly:peptide(s)).
XX
PS Claim 1; Col 139-144; 101pp; English.
XX
CC This sequence is encoded by a Porphyromonas gingivalis gene of the
CC invention. This sequence represents the hagA haemagglutinin protein. The
CC polypeptides are used to produce antibodies to organisms associated with
CC periodontal disease. The antibodies are also used in purification and
CC identification procedures. The genes and polypeptides are used as
CC vaccines against periodontal disease
XX
SQ Sequence 439 AA;

Query Match   95.0%;   Score 2321; DB 2;   Length 439;
Best Local Similarity   98.4%;   Pred. No. 5e-175;
Matches 426; Conservative   3; Mismatches   4; Indels   0; Gaps   0;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSSASYI 66

Qy   61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67 NFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126

Qy 121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db 127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIK 186

Qy 181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 246

Qy 241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306

Qy 301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYHYNCSDLNYILLDDIQFT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYHYNCSDLNYILLDDIQFT 366

Qy 361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 367 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNV 426

Qy 421 TVDPVQFNPVQNL 433
          :::||||||||||
Db 427 TINPTQFNPVQNL 439

RESULT 11
AAR96033
ID AAR96033 standard; protein; 1687 AA.
XX
AC AAR96033;
XX
DT 16-OCT-2003   (revised)

DT 04-SEP-1996   (first entry)
XX
DE P. gingivalis haemagglutinin hagE.
XX
KW Haemagglutinin; hagE; periodontal disease; vaccine; antibody.
XX
OS Porphyromonas gingivalis; strain FDC381.
XX
PN WO9617936-A2.
XX
PD 13-JUN-1996.
XX
PF 11-DEC-1995;   95WO-US016108.
XX
PR 09-DEC-1994;   94US-00353485.
XX
PA (UYFL) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
PI Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR WPI; 1996-287181/29.
DR N-PSDB; AAT30656.
XX
PT Porphyromonas gingivalis genes and proteins - used in the detection and
PT vaccination against periodontal disease.
XX
PS Claim 5; Page 138-143; 153pp; English.
XX
CC P. gingivalis 381 haemagglutinin hagE (AAR96033) was identified as the
CC product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
CC The haemagglutinin can be obtd. from transformed host cells and used as a
CC vaccine to protect humans or animals against periodontal disease.
CC Expression in Salmonella cells allows prodn. of a live vaccine. The
CC haemagglutinin can also be used to detect the presence of anti-P.
CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ Sequence 1687 AA;

Query Match   84.4%;   Score 2063; DB 2;   Length 1687;
Best Local Similarity   86.3%;   Pred. No. 8.9e-154;
Matches 389; Conservative   17; Mismatches   39; Indels   6; Gaps   4;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY- 59
          ||||||||||||||||||||||||||||||||||||||||||||||||:|||:
Db 939 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 995

Qy   60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          : |||||| |||||||||| ||||||||||||||||||||||||||||||||||||
Db 996 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1055

Qy 120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||| : |||| |||||||| :||||||:|| |||||||||||||||||||||||||
Db 1056 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1113

Qy 180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1114 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1173

Qy 240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1174 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1233

Qy 300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYHYNCSDLNYILLDDIQF 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1234 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYHYNCSDLNYILLDDIQF 1293

Qy 360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db 1294 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1353

Qy 420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450

```
Db   1354 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1384
          ||::|| |||||:||     |||:||
```

RESULT 12
AAW69495
ID   AAW69495 standard; protein; 1687 AA.
XX
AC   AAW69495;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagE.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagE.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995; 95US-00570311.
XX
PR   08-SEP-1988; 88US-00241640.
PR   25-JAN-1991; 91US-00647119.
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58881.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 167-182; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagE haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1687 AA;

Query Match   84.4%;   Score 2063; DB 2; Length 1687;
Best Local Similarity   86.3%;   Pred. No. 8.9e-154;
Matches   389; Conservative   17; Mismatches   39; Indels   6; Gaps   4;

Qy   1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY- 59
            ||||||||||||||||||||||||||||||| ||:|| :|
Db   939 PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 995

Qy   60 INFEGFQQNPDNYIVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          : |||||||||||| :|||||||::||||||||||||||||||||||||||| |||
Db   996 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1055

Qy   120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||| : |:|||| :||||||||| ||||||||||||||:||||:|| |:|:|:|:||
Db   1056 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLLDEVEI 1113

Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1114 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1173

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1174 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1233
```

```
Qy    300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1234 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1293

Qy    360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1294 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1353

Qy    420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  450
          ||::| |||||:|:|| | ||||||||||:|:
Db   1354 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1384
```

RESULT 13
AAR70188
ID   AAR70188 standard; protein; 1704 AA.
XX
AC   AAR70188;
XX
DT   25-MAR-2003 (revised)
DT   21-SEP-1995 (first entry)
XX
DE   Arg-gingipain-2 prepolyprotein.
XX
KW   Arg-gingipain-2; gingivalis; periodontal disease; vaccine;
KW   arginine-specific protease.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key   Location/Qualifiers
FT   Protein   228..719
FT   /label= "Protease
FT   /note= "corresponds to Arg-gingipain-1"
FT   Region   720..1091
FT   /label= Hemagglutinin
FT   Region   1092..1429
FT   /label= Hemagglutinin
FT   Region   1430..1704
FT   /label= Hemagglutinin
XX
PN   WO9507286-A1.
XX
PD   16-MAR-1995.
XX
PF   09-SEP-1994; 94WO-US010283.
XX
PR   10-SEP-1993; 93US-00119361.
PR   21-OCT-1993; 93US-00141324.
PR   24-JUN-1994; 94US-00265441.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J, Potempa J, Barr PJ, Pavloff N;
XX
DR   WPI; 1995-123373/16.
DR   N-PSDB; AAQ83489.
XX
PT   DNA encoding Arg-gingipain proteins - used to develop prods. for
PT   detection, treatment and prevention of periodontal disease.
XX
PS   Disclosure; Page 70-77; 89pp; English.
XX
CC   A low mol.wt. arginine-specific gingipain (AG-1) and high mol.wt. AG (AG-
CC   2) were isolated from P. gingivalis strains H66 (ATCC 33277) and W50
CC   (ATCC 53973). The sequences of the proteins were used to design PCR
CC   primers and probes to isolate AG DNA. Lambda DASH and lambda ZAP
CC   libraries were screened with a probe based on amino acids 11-22 of the AG
CC   protein to obtain DNA encoding AG-1 (AAQ83484) and AG-2 (AAQ83489). AG-2
CC   is a prepolyprotein incorporating AG-1. (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 1704 AA;

```
Query Match       84.4%;   Score 2063;   DB 2;   Length 1704;
Best Local Similarity   86.3%;   Pred. No. 9e-154;
Matches 389;   Conservative 17;   Mismatches 39;   Indels 6;   Gaps 4;

Qy    1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-   59
        ||||||||||||||||||||||||||||||:        ||:|| : :|
Db  956 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1012

Qy   60 INFEGPQNPDNVLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
         :     ||||:||:|||||||||:|||||||||||||||||||||||||||||||:|||
Db 1013 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1072

Qy  120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
        ||  :|||| :  |||||| |||||:||||:|||||||||||||  ::|:|:|| ::|||
Db 1073 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1130

Qy  180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF  239
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1131 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1190

Qy  240 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  299
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1191 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1250

Qy  300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
        ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db 1251 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1310

Qy  360 TMGGSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1311 TMGGSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1370

Qy  420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
        ||::|:|||||:||||:  :||| ||:
Db 1371 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

RESULT 14
AAW34843
ID AAW34843 standard; protein; 1704 AA.
XX
AC AAW34843;
XX
DT 03-JUN-1998 (first entry)
XX
DE Arg-gingipain high molecular weight prepolyprotein sequence.
XX
KW Arg-specific gingipain protease; gingivalis; periodontal disease;
KW vaccine; infection.
XX
OS Porphyromonas gingivalis.
XX
FH Key   Location/Qualifiers
FT Protein   1..227
FT   /note= "precursor protein"
XX
PN WO9734629-A1.
XX
PD 25-SEP-1997.
XX
PF 21-MAR-1997; 97WO-US004635.
XX
PR 22-MAR-1996; 96US-0013945P.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
PA (MORE-) MOREHOUSE SCHOOL MEDICINE.
XX
PI Potempa J, Travis J, Genco C;
XX
DR WPI; 1997-479993/44.
DR N-PSDB; AAT93872.
XX

PT Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
PT useful for protecting animals and humans from gingivalis and periodontal
PT diseases.
XX
PS Disclosure; Page 68-73; 95pp; English.
XX
CC The present sequence represents an arginine-specific protease of
CC Porphyromonas gingivalis. The following peptides, derived from Arg- and
CC Lys-specific high molecular weight proteases, offer protection against
CC infection: YTYTVRDGK IKEGLTATTE DDGVATGNHE YCVEKYTAGS VSPKVC (I);
CC YTPVEEKQNG RMVIVAKKY (II); QLPPIFDVAC VNGDFLFSMP CFAEALMRAQ (III);
CC GEPNPYQPVS NLTATTQGQK VTLKWDAPSTK (IV); GNHEYCVEVK YTAGVSPKVC KDVTV (V);
CC RMFMNYEPGR YTPVEEKQNG (VI); TFAGFEDTYK RMFMNYEPGR (VII); DYTYTVYRDG
CC TKIKEGLTAT TFEEDGVATG NMEYCVCVKY TAGVSPKVC (VIII); YTYTVYRDGT KIKEGLTATTF
CC EEDG (IX); RDGTKIKEGL TATTFEEDGV ATGN (X); KIKEGLTATT FEEDGVATGN HEY (XI)
CC ; KWDAPNGTPN PNPN PGTTLSE (XII); and YTPVEEKENG RMIVIVAKKY
CC (XIII). They are used in vaccines to protect animals, including humans,
CC from gingivitis and/or periodontal diseases
XX
SQ Sequence 1704 AA;

```
Query Match       84.4%;   Score 2063;   DB 2;   Length 1704;
Best Local Similarity   86.3%;   Pred. No. 9e-154;
Matches 389;   Conservative 17;   Mismatches 39;   Indels 6;   Gaps 4;

Qy    1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-   59
        ||||||||||||||||||||||||||||||:        ||:|| : :|
Db  956 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1012

Qy   60 INFEGPQNPDNVLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
         :     ||||:||:|||||||||:|||||||||||||||||||||||||||||||:|||
Db 1013 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1072

Qy  120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
        ||  :|||| :  |||||| |||||:||||:|||||||||||||  ::|:|:|| ::|||
Db 1073 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1130

Qy  180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF  239
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1131 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1190

Qy  240 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  299
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1191 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1250

Qy  300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
        ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db 1251 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1310

Qy  360 TMGGSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1311 TMGGSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1370

Qy  420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
        ||::|:|||||:||||:  :||| ||:
Db 1371 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

RESULT 15
AAY67396
ID AAY67396 standard; protein; 1704 AA.
XX
AC AAY67396;
XX
DT 25-APR-2000 (first entry)
XX
DE Arg-gingipain-2 amino acid sequence.
XX
KW Arginine specific proteinase; Arg-gingipain; gingipain-2; haemagglutinin;
KW immunogenic component; vaccine; inflammatory response; tissue damage;
KW periodontal disease.
XX
OS Porphyromonas gingivalis.

```
Qy    420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          ||::||||||||||        ||  ||  ||:|||
Db   1371 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401

Search completed: May 18, 2004, 11:42:40
Job time : 47.4099 secs
```

```
XX
FH   Key        Location/Qualifiers
FT   Region     229..719
FT              /note="Amino acids 229-719 are specifically claimed"
FT   Region     720..1185
FT              /note="Amino acids 720-1185 are specifically claimed"
XX
PN   US6017532-A.
XX
PD   25-JAN-2000.
XX
PF   08-NOV-1994;   94US-00336308.
XX
PR   10-SEP-1993;   93US-00119361.
PR   24-JUN-1994;   94US-0265441.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Potempa JS, Travis J;
XX
DR   WPI; 2000-136659/12.
DR   N-PSDB; AAZ60181.
XX
PT   New Porphyromonas gingivalis arginine-specific protease preparation
PT   useful for preparing vaccines against periodontal disease and for
PT   screening for Arg-gingipain inhibitors.
XX
PS   Claim 1; Col 29-42; 55pp; English.
XX
CC   This sequence represents a Porphyromonas gingivalis arginine-specific
CC   proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
CC   Gingipain-2 consists or a 50kD protease component non-covalently
CC   associated with a 44kD haemagglutinin component. The proteinase is
CC   stimulated by glycine containing peptides and glycine analogues. It is
CC   inhibited by cysteine protease group specific inhibitors. The protease
CC   preparation can be used in immunogenic compositions and vaccines against
CC   inflammatory response and tissue damage caused by P. gingivalis in
CC   periodontal disease. It can also be used to screen for agents that
CC   modulate Arg-gingipain proteinase activity inhibitors
XX
SQ   Sequence 1704 AA;

Query Match            84.4%;     Score 2063; DB 3;  Length 1704;
Best Local Similarity  86.3%;     Pred. No. 9e-154;
Matches 389;  Conservative  17;  Mismatches  39;  Indels   6;  Gaps     4;

Qy      1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASY- 59
          ||||||||||||||||||||||||||||||||         ||||||: |: :
Db    956 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSRSFG 1012

Qy     60 INFEGPQNPDNYLIVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          : |||||||||||: ||  ||||||||||||||||||||||||||||||||
Db   1013 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1072

Qy    120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||| ||| : :| |||||||||||||||||||||||||||||||||: |: :|||||||
Db   1073 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1130

Qy    180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1131 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1190

Qy    240 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFBET 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1191 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFBET 1250

Qy    300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1251 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1310

Qy    360 TMGGSPTPPDYTYTVYRDGTKIIKEGLTETTFBEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1311 TMGGSPTPPDYTYTVYRDGTKIIKEGLTETTFBEDGVATGNHEYCVEVKYTAGVSPKECVN 1370
```

```
            GenCore version 5.1.6
       Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     May 18, 2004, 11:37:00 ; Search time 11.7493 Seconds
                              (without alignments)
                        3684.135 Million cell updates/sec

Title:          US-08-570-311-16
Perfect score:  2443
Sequence:       1 FNPNPGTTTLSESFENGIPA.........QNLTGSAVGQKVTLKWDAPN 450

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:    283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      PIR_78:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2443	100.0	2628	2	T28651	hemagglutinin A -
2	2063	84.4	1704	2	A55426	gingipain R (EC 3.
3	2040	83.5	1526	2	S49763	gingipain R (EC 3.
4	2011.5	82.3	1732	2	T30836	lysine-specific cy
5	266	10.9	991	2	I40229	arginyl endopeptid
6	132.5	5.4	1052	2	AF2959	conserved hypothet
7	132.5	5.4	1341	2	H98323	hypothetical prote
8	130.5	5.3	2468	2	A83412	hypothetical prote
9	127.5	5.2	1684	2	S10789	amylase A-180 - al
10	126	5.2	1649	2	C86822	hypothetical prote
11	125.5	5.1	5291	2	F90696	hypothetical prote
12	124.5	5.1	691	2	B75622	hypothetical prote
13	124.5	5.1	2817	2	B97033	uncharacterized pr
14	124	5.1	5188	2	B85547	probable RTX famil
15	123.5	5.1	1904	2	T13256	tail-host specific
16	122	5.0	713	2	B75489	hypothetical prote
17	121.5	5.0	465	2	A47023	S-layer protein -
18	121	5.0	1090	2	S59077	cellulose 1,4-beta
19	121	5.0	4936	2	AH2515	hypothetical prote
20	120	4.9	908	2	AE2254	hypothetical prote
21	120	4.9	1345	2	H90975	hypothetical prote
22	120	4.9	1873	2	T30944	surface protein pr
23	120	4.9	2783	2	T34416	hypothetical prote
24	120	4.9	3624	2	AD0835	large repetitive p
25	119	4.9	1441	2	A86685	prophage pil prote
26	119	4.9	1461	2	E90696	hypothetical prote
27	119	4.9	1461	2	A85547	hypothetical prote
28	118.5	4.9	926	2	D86897	hypothetical prote
29	118.5	4.9	1939	2	D97316	probable S-layer p
30		116.5		4.8	1180	2	E86719	hypothetical prote
31		116.5		4.8	1385	2	T18213	parasporal crystal
32		116		4.7	2660	2	E85822	probable invasin Z
33		116		4.7	3029	2	S76109	hypothetical prote
34		115.5		4.7	715	2	JC4908	alkaline serine pr
35		115.5		4.7	1034	2	T30551	beta-galactosidase
36		115.5		4.7	1118	1	A49724	protein-tyrosine-p
37		115		4.7	821	2	AD1507	probable secreted
38		114.5		4.7	607	2	C69503	conserved hypothet
39		114.5		4.7	1248	2	C89974	autolysin [importe
40		113.5		4.6	875	2	AF0472	probable outer mem
41		113.5		4.6	938	2	AF1772	internalin-like pr
42		112.5		4.6	987	2	A64474	hypothetical prote
43		111.5		4.6	1122	2	T18346	MGC1 protein precu
44		111.5		4.6	1274	2	T10729	transferrin-like p
45		111.5		4.6	2167	2	AF1489	cell wall-associat

ALIGNMENTS

```
RESULT 1
T28651
hemagglutinin A - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C;Accession: T28651
R;Han, N.; Whitlock, J.; Progulske-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A;Title: The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381 contains four
A;Reference number: Z20494; MUID:97047672; PMID:8926061
A;Accession: T28651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2628 <HAN>
A;Cross-references: EMBL:U41807; NID:g1552410; PID:g1469916; PIDN:AAB17128.1
C;Genetics:
A;Gene: hagA

Query Match             100.0%;  Score 2443;  DB 2;  Length 2628;
Best Local Similarity   100.0%;  Pred. No. 8.7e-161;
Matches  450;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI  60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       500 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 559

Qy        61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       560 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 619

Qy       121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       620 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 679

Qy       181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       680 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 739

Qy       241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       740 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 799

Qy       301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       800 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 859

Qy       361 MGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       860 MGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 919

Qy       421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
             ||||||||||||||||||||||||||||||
```

Db   920 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 949

RESULT 2
A55426
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
N;Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP; R
C;Species: Porphyromonas gingivalis
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
C;Accession: A55426; D53113
R;Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr, P
J. Biol. Chem. 270, 1007-1010, 1995
A;Title: Molecular cloning and structural characterization of the Arg-gingipain proteina
A;Reference number: A55426; MUID:95138080; PMID:7836351
A;Accession: A55426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1704 <PAV>
A;Cross-references: GB:U15282; NID:g557067; PIDN:AAA69539.1; PID:g557068
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: D53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 228-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141694)
C;Keywords: cysteine proteinase; hydrolase

Query Match   84.4%;   Score 2063;   DB 2;   Length 1704;
Best Local Similarity   86.3%;   Pred. No. 1.1e-134;
Matches   389;   Conservative   17;   Mismatches   39;   Indels   6;   Gaps   4;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNWTTTPPGGTSFAGHNSAICASSASY-   59
           ||||||||||||||||||||||||||||||||||| ||||:|| | |: | :
Db   956 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1012

Qy   60 INFEGPQNPDNYLIVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
           || | |:|||||:|| || |||| ||||||||||||||||||||||||||||||||
Db   1013 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1072

Qy   120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
           |||:||| |:||||||||||||||||| | :||||||||||||||| || ::|||||
Db   1073 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1130

Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1131 KANGKRADFTETFESSTHGEATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1190

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEET 299
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1191 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEET 1250

Qy   300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1251 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1310

Qy   360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1311 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1370

Qy   420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
           |:|| ||||||:|| :| |||||||||||
Db   1371 VTINPTQFNPVQNLKAQPDGGDVVLKWEAPS 1401

RESULT 3
S49763
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997
A;Molecule type: DNA

C;Accession: S49763
R;Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.
submitted to the EMBL Data Library, November 1994
A;Description: Cloning, sequence analysis and expression in Escherichia coli of prpRl of
A;Reference number: S49763
A;Accession: S49763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <ADU>
A;Cross-references: EMBL:X82680
C;Genetics:
A;Gene: prpR1
C;Keywords: cysteine proteinase; hydrolase

Query Match   83.5%;   Score 2040;   DB 2;   Length 1526;
Best Local Similarity   85.6%;   Pred. No. 3.6e-133;
Matches   386;   Conservative   17;   Mismatches   42;   Indels   6;   Gaps   4;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNWTTTPPGGTSFAGHNSAICASSASY-   59
           ||||||||||||||||||||||||||||||||||| ||||:|| | |: | :
Db   958 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1014

Qy   60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
           || | |:|||||:|| || |||| ||||||||||||||||||||||||||||||||
Db   1015 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1074

Qy   120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
           |||:||| |:||||||||||||||||| | :||||||||||||||| || ::|||||
Db   1075 LEETITAKG-VRSPEAIRG-RIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1132

Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133 KANGKRADFTETFESSTHGEATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSF 1192

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEET 299
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEET 1252

Qy   300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1253 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1312

Qy   360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1313 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVN 1372

Qy   420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
           |:|| ||||||:|| :| |||||||||||
Db   1373 VTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 4
T30836
lysine-specific cysteine proteinase porphypain (EC 3.4.22.-) - Porphyromonas gingivalis
N;Alternate names: lysine-specific cysteine proteinase 1, 60K
C;Species: Porphyromonas gingivalis
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30836; T30837; T30526; A53113
R;Barkocy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulske-Fox, A.; Lantz
J. Bacteriol. 178, 2734-2741, 1996
A;Title: Analysis of the prtP gene encoding porphypain, a cysteine proteinase of Porphy:
A;Reference number: Z20895; MUID:96213011; PMID:8631659
A;Accession: T30836
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1732 <BAR>
A;Cross-references: EMBL:U42210; NID:g1314325; PID:g1314326; PIDN:AAB06565.1
R;Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z20896
A;Accession: T30837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-991 <RES>
A;Cross-references: GB:D26470; NID:g927644; PIDN:BAA05484.1; PID:g927645

Query Match   10.9%;   Score 266; DB 2;   Length 991;
Best Local Similarity   25.9%;   Pred. No. 1.7e-10;
Matches 114; Conservative 50; Mismatches 164; Indels 112; Gaps 19;

```
Qy   24 TIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYINFEGPQNPDNYLVTPELSLPNGG   83
                          :::        |||  :: :    :  |:    :|:
Db  595 SVNVSCDYNGAIATISANGKMF---GSAVVENGTATINLTG---------------LTNES  637

Qy   84 TLTFWVCAQDANYASEHYAVYASSTG--NDASNFANALLEEVLTAKTVVTAPEAIRGTRV  141
                |||                 :|:|:          ::     | :| :|:    :|
Db  638 TLTLTV----VGYNKETVIKTINTNGEPNPYQPVSN-----LTATT-------QGQKV  679

Qy  142 QGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV-EIKANGKRADFTETFSSTHGEA  200
               |||       ::||||||||||||||     ||||| |:  |||||||||||:||::
Db  680 TLKWDAPSTKTNATTN-TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWNDGS-  737

Qy  201 PAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWN-----GMALNPDNYLISK  255
               |    |||||||| : |:   ::   |:||:|||:|: :|:  : ...||:||||||||:
Db  738 -GYQILLDADHDQYGQVIPSDIHTLWPNCSVPANLFAPFEYTVPENADPSCSPTNMIM--  794

Qy  256 DVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTE  315
               :|:        ||:                            ::              ||
Db  795 DGTASVNI------------------------PAGTY-------------------DFAI  811

Qy  316 ADGAKPQS--VWI------ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP  365
               .:     ::||        :        |||  :|||  ||||:|||
Db  812 ---AAPQANAKIWIAGQGPTKEDDYVFEAGKKY---HFLMKKMGSGDGTELTISEGGG-  863

Qy  366 TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDFPV  425
               :.|||:|||||||||||||    :|::|||:|||||||||||||||:||||:
Db  864 --SDYTTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVEVKYTAAGVSPKVCVDYIPDGV  921

Qy  426 QFNPVQN-LTGSAVGQKVTL  444
               |||.| |  ::|:|:
Db  922 ADVTAQKPYTLTVVGKTITV  941
```

RESULT 6
AF2959
conserved hypothetical protein Atu3276 [imported] - Agrobacterium tumefaciens (strain C5...
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 18-Nov-2002
C;Accession: AF2959
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I...
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell...
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kreppan, W.; Perry, M.; Gordon-Kamm,...
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF2959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1052 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44092.1; PID:g17741659; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3276
A;Map position: linear chromosome

Query Match   5.4%;   Score 132.5; DB 2;   Length 1052;
Best Local Similarity   20.7%;   Pred. No. 0.32;
Matches 122; Conservative 67; Mismatches 160; Indels 239; Gaps 33;

```
Qy    7 TTTLSESFENGIPASWKTIDADGDGNNWTI---------TPPPGGTSFAGHNSAICASSAS   58
                 ||  |       |  ::             |    ||||       ||    :|
Db   52 TVTSGEAIIGRHSASRATV--TGDGSKWTTGDLQVGGDTSDPGG--LAGNGT--------   99
```

A;Residues: 1-795,'I',797-1389,'N',1391-1478,'Y',1480-1732 <SLA>
A;Cross-references: EMBL:U75366; NID:g2182811; PID:g2182812; PIDN:AAB60809.1
R;Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A;Title: IS195, an insertion sequence-like element associated with protease genes in Por...
A;Reference number: Z20844; MUID:9829816; PMID:9632563
A;Accession: T30526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1350,'N',1352-1363,'Y',1365-1447,'H',1449-1732 <LEW>
A;Cross-references: EMBL:AF017059; NID:g2738802; PID:g2738803; PIDN:AAC26523.1
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat...
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: A53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 229-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141690)
C;Genetics:
A;Gene: prtP; prtK
C;Keywords: cysteine proteinase; hydrolase

Query Match   82.3%;   Score 2011.5; DB 2;   Length 1732;
Best Local Similarity   84.7%;   Pred. No. 4e-131;
Matches 383; Conservative 19; Mismatches 41; Indels 9; Gaps 6;

```
Qy    1 PNPNDGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-   59
                |||| |||||||||||||||||||||||||||||   ||||||||||||||||||||
Db  977 PNPNFG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1032

Qy   60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
                ||    || ||| :||  | |  ||||||||||||||||||||||||||||||||||||
Db 1033 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1092

Qy  120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
               |||:|||| :|  :|| ||| :|||| ||||||||||||||||||||||||||||||
Db 1093 LEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1150

Qy  180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF  239
               |||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db 1151 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVASF 1210

Qy  240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  299
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1211 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1270

Qy  300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
               ||||||||||||||||| :||||||||||||||||||||||||||||||||||||||||
Db 1271 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1330

Qy  360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db 1331 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVD 1390

Qy  420 VTVDPVQFNFVQNLTGSAV-GQKVTLKWDAP  449
               ||| ||| |||||| ||   |:||||  ||||
Db 1391 VTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422
```

RESULT 5
I40229
arginyl endopeptidase - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40229
R;Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A;Title: Structural characterization of argingipain, a novel arginine-specific cysteine...
A;Reference number: I40229; MUID:95168884; PMID:764651
A;Accession: I40229

```
Qy   59 YINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGN-------- 110
Db  100 ------------LNVTAGGSV------DSTVA--HLGVVAGATGSAIVDGKGS 132
Qy  111 ---DASNF-----ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVA 160
Db  133 VWTVDRNSLEVGVSGAGSL--AVTGGGLVDAANIIIGTNTGG---NGSVRVSGADSTVK 186
Qy  161 FRHFGCTDFFWINL-----DDVEIKANG---KRADFTETFESSTHGEAPAEWTTIDADGD 212
Db  187 SR-----SD---LNVGLYGNGSMTVEAGGAVKSRDGYVATYGGST----------SAVTVTGD 231
Qy  213 GQGWLCLSSGQLDWLTAHGGT-NVVASFSWNGMALNPDNYLI----------SKDVTGA-TK 262
Db  232 GSSWAMTGTFFVGY--ASGATGNVTVS--NGGAIRATGVTLGDLAGASGTMTITGAGSK 286
Qy  263 VKYYY---------------DHY----------AVMISKTGT--- 287
Db  287 VTAYVDNGTVNSGSVDVGFQGSGSLSVVNGGSLDAYNLYVGNALGSSGAVLVSGVGSHVS 346
Qy  288 ------NAGD----------FTVVFEETPNGINKGGARFGLSTEADGAKPQSV 324
Db  347 VDGLMVVGNAGNGSVEITGGASLAAPTLIATEAGSTGVLSIGAGSGQTARSAGA------ 401
Qy  325 WIB-RTVDLPAGTKYVAFRHYNCSDLNYILLDDIQ------FTMGGSPTPTDYTY 372
Db  402 -VEARAIAFGAGNGSIVFNH--SETGYTLSADISGAGRVVAEAGVTTLSGNNS- 451
Qy  373 TVYRDGTKIKEGLTETT----------------FEEDG----VA 396
Db  452 --YSGGTTISAGMLKGTAKSFGSGGIVNNAELVVDGGGTLSNAISGTGSFEKTGDGNLLL 509
Qy  397 TGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLTGSAVGQKVTL 444
Db  510 TGNSTYSGATAVSAG-----KLSVNGSLASAVSVGSGATVGGTGTI 550
```

RESULT 7
H98323
hypothetical protein AGR_L_3085 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: H98323
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: H98323
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1341 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK90114.1; PID:g15160106; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_3085
A;Map position: linear chromosome

Query Match   5.4%; Score 132.5; DB 2; Length 1341;
Best Local Similarity 20.7%; Pred. No. 0.45;
Matches 122; Conservative 67; Mismatches 160; Indels 239; Gaps 33;

```
Qy    7 TTTLSESFENGIPASWKTIDADGDGNNWTT----------TPPPGGTSFAGHNSAICASSAS 58
Db  341 TVTSGERAIIGRHSASEATV--TGDGSKWTTGDLQVGGDTSDPGG--LAGNGT-------- 388
Qy   59 YINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGN-------- 110
Db  389 ------------LNVTAGGSV------DSTVA--HLGVVAGATGSAIVDGKGS 421
Qy  111 ---DASNF-----ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVA 160
Db  422 VWTVDRNSLEVGVSGAGSL--AVTGGGLVDAANIIIGTNTGG---NGSVRVSGADSTVK 475
```

```
Qy  161 FRHFGCTDFFWINL-----DDVEIKANG---KRADFTETFESSTHGEAPAEWTTIDADGD 212
Db  476 SR-----SD---LNVGLYGNGSMTVEAGGAVKSRDGYVATYGGST----------SAVTVTGD 520
Qy  213 GQGWLCLSSGQLDWLTAHGGT-NVVASFSWNGMALNPDNYLI----------SKDVTGA-TK 262
Db  521 GSSWAMTGTFFVGY--ASGATGNVTVS--NGGAIRATGVTLGDLAGASGTMTITGAGSK 575
Qy  263 VKYYY---------------DHY----------AVMISKTGT--- 287
Db  576 VTAYVDNGTVNSGSVDVGFQGSGSLSVVNGGSLDAYNLYVGNALGSSGAVLVSGVGSHVS 635
Qy  288 ------NAGD----------FTVVFEETPNGINKGGARFGLSTEADGAKPQSV 324
Db  636 VDGLMVVGNAGNGSVEITGGASLAAPTLIATEAGSTGVLSIGAGSGQTARSAGA------ 690
Qy  325 WIB-RTVDLPAGTKYVAFRHYNCSDLNYILLDDIQ------FTMGGSPTPTDYTY 372
Db  691 -VEARAIAFGAGNGSIVFNH--SETGYTLSADISGAGRVVAEAGVTTLSGNNS- 740
Qy  373 TVYRDGTKIKEGLTETT----------------FEEDG----VA 396
Db  741 --YSGGTTISAGMLKGTAKSFGSGGIVNNAELVVDGGGTLSNAISGTGSFEKTGDGNLLL 798
Qy  397 TGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLTGSAVGQKVTL 444
Db  799 TGNSTYSGATAVSAG-----KLSVNGSLASAVSVGSGATVGGTGTI 839
```

RESULT 8
A83412
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83412
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83412
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2468 <STO>
A;Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN00
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA1874

Query Match   5.3%; Score 130.5; DB 2; Length 2468;
Best Local Similarity 22.2%; Pred. No. 1.4;
Matches 119; Conservative 47; Mismatches 182; Indels 189; Gaps 30;

```
Qy    2 NPNPGTT----------TLSESFENGIPASWKTIDADGDGN---------NWT--- 35
Db  757 NPSNGTTLSGTAEPGSSVTLTDG--NGNPIG--QVTADGSGNWSFTPSTPLADGTVVNAT 812
Qy   36 TTPPPGGTSFAGHN----------SAICASSASYINFEGPQNP---------- 68
Db  813 AIDPAGNTSGQGSTTVDGVAPTTPTVNLSNGSSLSGTAEPGSTVILTDGNGNPIAEVTAD 872
Qy   69 --DNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDASNFANALLEEVLTA 126
Db  873 GSSGNWTVTPSTPIANGTVVN--VVAQDA---------AGN--SSPGASVTVDSQAPA 916
Qy  127 KTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGKRA 186
Db  917 APVVNPS--NGTTLSGT---------------VTLTD---GNGNPI 948
Qy  187 DFTETFESSTHGEAPAEWTTIDADGDGQGW----------LCLSSGQLDWLTAH---GGTNVVAS 238
```

```
Db    949  G-----------QVTADGSG-NWSFTPGTPLANGTVVNATASDPTGNTSAPAS 989
               |
Qy    239  FSWNGMA-----LNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTG----- 286
             :|| |||::|||
Db    990  TTVDSVAPAAPVVNPSN--GAEISGTAEPGATVTLTDGSGNPIGQVTADGSGNWSFTPS 1046
             ||| :||
Qy    287  TNAGDFTVV-FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHY 344
             || : :
Db   1047  TPLADGTVVNATATDPAG-NTGGQG---STTVDAIAPAT---PTVNLSNGSSLSG---- 1094
Qy    345  NCSDLNYILLDD------IQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATG 398
                             |||
Db   1095  TAEPGSTVILTDGNGNPIAEVTADGS---GNWTYT---PSTPIANGTVVNVAQD--ASG 1146
               :|| ||
Qy    399  NHEYCVEVKYTAGVSPKECVN----------VTVDPVQFNPVQNLTGSAVG 439
             || :||
Db   1147  NSSPPATVTVDSSAPAPVINPSNGVVISGTAEAGATVTLTDAGGNPIGQVTADGSG 1203

RESULT 9
S10789
amylase A-180 - alkaliphilic eubacterium 163-26
C;Species: alkaliphilic eubacterium 163-26
C;Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 15-Oct-1999
C;Accession: S10789
R;Candussio, A.; Schmid, G.; Boeck, A.
Eur. J. Biochem. 191, 177-185, 1990
A;Title: Biochemical and genetic analysis of a maltopentaose-producing amylase from an a
A;Reference number: S10789; MUID:9336627; PMID:1696201
A;Accession: S10789
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1684 <CAN>
A;Cross-references: EMBL:X53373; NID:g48305; PIDN:CAA37453.1; PID:g48306

Query Match            5.2%;    Score 127.5;   DB 2;    Length 1684;
Best Local Similarity 20.9%;    Pred. No. 1.3;
Matches  102;  Conservative  53;  Mismatches 184;  Indels 149;  Gaps  27;

Qy     17  GIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA-SSASYINFEGPQNPDNYLVTP 75
               :|   | : :
Db    204  GLPRDWTPNQA--QGQNWHT------------HNDIMNKDNEAAWANWWGSD-------- 241
Qy     76  ELSLPNGGTLTFWVCAQD-ANYASEHYAVYASSTGNDASNFANALLEEVLTAKTV-VTAP 133
                 : : :|
Db    242  ------------DNCGGSEQTMCIGFLPDIKTEVTTGVDLP 280
Qy    134  EAIRGT-RVQGTWYQKTVQLPAGTKYVAFRHFGCTDFF--WINLDDVEIKANGKRADFTE 190
                      |  ||
Db    281  PILRNKWNDQASGYEDWF-VPAAEPYRQDLNIAPKDYLIKWITSWVEEFGIDGFRVDTAK 339
Qy    191  TFESSTHGEAPAE---------WTTIDADGDGQGWLCLSSGQLDWLTA--------HG 231
Db    340  HVEIERWAELKNEAEVALQTWRENNPDKPGANW-----DDNFWMTAEVFGHGLGKSEYFD 394
Qy    232  -GTNVVASFSW----------NGMALNPDNYLISKDVTGATKVKYYYAVNDGFP 274
                                   ||: ||
Db    395  FGFDSVINFEFQNANFNNLEGLFSRYANSINTDPDFNMLSYVSSHDTKL---YSRDD--- 448
Qy    275  GDHYAVMISKTGTN---AGDFTVVF-EETPNGINKGGARFGLSTEADGAKPQSVW--IE 327
Db    449  ------LIQAGTALLLLPGGVQVFYGDETARPLGDGG----SDPEQGTRSSMNWANIN 496
Qy    328  RTVDLPAGTKYVAFR------------HYNCSDLNYI---------LLDDIQFTMGGSPTPT 368
             :||  : |
Db    497  QNV-LSHWQKLGQFRNNHIAIGAGAHQKLSDSPYYFARITYESDDIVDEVVVATGAQGTTA 555
Qy    369  DYTYTVYRDGTKIKEGLT--ETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPVQ 426
Db    556  VTVBGVFEDGTVVRDAYTGDETTVTK-GTAT----------FTAGTQGIILIENTAEPVT 604
Qy    427  FNPVQNLT 434
             :|:
```

```
Db    605  NLPIVSAT 612

RESULT 10
C86822
hypothetical protein yqbK [imported] - Lactococcus lactis subsp. lactis (strain Il1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: C86822
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: C86822
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1649 <STO>
A;Cross-references: GB:AE005176; PID:g12724583; PIDN:AAK05677.1; GSPDB:GN00146
A;Experimental source: strain Il1403
C;Genetics:
A;Gene: yqbK

Query Match            5.2%;    Score 126;   DB 2;    Length 1649;
Best Local Similarity 20.4%;    Pred. No. 1.6;
Matches  110;  Conservative  56;  Mismatches 189;  Indels 184;  Gaps  28;

Qy     40  PGGTSFAGHNSAICASSASYINFEG-PQN--------LVTPELSLPN 81
             |||||  : :|||||                     :
Db     65  PSGTTFNG-------NSVSQPNRTPLPQNSVEDAIMPNSFGWQQTIANLTLPTNQVTAVQ 117
Qy     82  GGTLTFWVCAQDANYASEHYAVYASSTGNDASNF-----ANALLEEVLTAKTVVTAPRA- 135
             ||: |
Db    118  NGTIIGSYPTTDAGVQQALYAMY-SATNASTSDFGLYYGANTTLSNSL-LKAVVSNFSAT 175
Qy    136  ------IRGTRVQGTWYQKTVQL----------PAGTKYVAFR----HFGCTDFFWINLDDV 177
              : |||  || :|:|                  :||
Db    176  NMTFSTLKGHAKSLTWISNPADLLTSNSSQPSGTNYTGTLPSNVYFGVPTIF----RNV 231
Qy    178  EIKANGKRADFTETFESSTHGEAPAEWTT------------IDADGDGQGW 216
              : |||                                   ::| : |
Db    232  TVAASGDNIYAQGNAFATTNG---SWITGAPNIYGGTDNSDISGNTNLYIGATGSIAGW 287
Qy    217  LCLSSGQLDWLTAHGGTNV-------------VASFSWNGMALNPDNYLISKDVTG--ATKV 263
              :|
Db    288  -NIYGGNASAATISGNTHVTIAQSSTINSVTGGSASGTTISGNTNJ--DISGAIASQI 343
Qy    264  KYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNG-----INKGGARFG-----LS 313
              |: ||
Db    344  TNIYGAGIG-------TSNSPVNVNGNVTTYVNSTNGGARYQLYQGGTVYGNISGSIY 394
Qy    314  TEADGAKPQSWIERTVDJ----------PA----------GTKYVAFRHYNCSDLNYIL 353
              :||
Db    395  NTLSGA---GGWTGATSNINGAGGPASTFNGGSFQGNIGTSGAGNVISNSYNTSSFT--- 448
Qy    354  LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG-- 411
              :||
Db    449  TGQALFTGGNAGTGASYAQA--------TNSTTAAQGILYANITNYIKSAFTTGTA 496
Qy    412  -------------VSPKE----CVNVTVDPVQFNPVQNLTGSAVGQKVT 443
              : |
Db    497  GAVYGIVGGNGHDSLKISPSOWGLGSQTGLDSAVGVT-DAKAYGQIPSTTVVSNAQKIT 554

RESULT 11
F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: F90696
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
```

```
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90696
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5291 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33965.1; PID:g13360000; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0542

Query Match       5.1%;   Score 125.5;  DB 2;  Length 5291;
Best Local Similarity   22.2%;   Pred. No. 8.1;
Matches  113;  Conservative  58;  Mismatches 165;  Indels 173;  Gaps 28;

Qy   4 NPGTTTLSESFE---------NGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA  54
         ||| |||||                  |||||||||||||||     :   | | ::
Db 361 NPLTISGGSTAEAGQTVTVTLNGVTYS-GSVQADG---SWSVSLPTADL------SNLTA 410

Qy  55 S----SASYINFEGPQNPDNYLVTPELSLPNGGTLTF-WVCAQDANYASEH---YAVVAS 106
           :::: |   ||:   |       | | :||   : :    :
Db 411 SQYTVSASVSDKAGNPASANHGLAVDLTVP---VLTINTVSGDDIINAAEHGQALVISGS 467

Qy 107 STGNDASNFANALLEEVLTAKTVTVTAPEAIRGTRVQGTWYQKTVQLPAG----------155
         |||  |    :::::|   :||                |
Db 468 STGGEAGD---VITVTLNSKTYTTMLDA------SGNW---SVGVPAADVTALGSGPQT 514

Qy 156 ---------------TKYVAFRHFGCTDFFWINLDDVEIKANGKRADFTETFPESST 196
                          ::|                :||||||||||||||
Db 515 ITAAITDAAGNSDDASRTVTVNLAAPTIGINT---IATDDV-IKATEKGADLQITGTSN- 569

Qy 197 HGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKD 256
           | | :                           :  :    ::
Db 570 --QPA-GTTITVTLNGQNYTATTDSNGNW----SATVPASAV--SALGEANYTVTAN    617

Qy 257 VT----GATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGL 312
         |         :                             ::
Db 618 VTDTAGNSNSASHNVLVNSALP-----AVTINAVAT------DDIINAAESGNAQ-TI   663

Qy 313 STEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTY 372
            :                     ::        :          ||
Db 664 SGQVTGAAQG------------------DIVTVTLGGN-----TY              685

Qy 373 TVYRDGTKIKEGLTETTF--EEDGVATGNHEYCVEVKYTAGV--SPKECVNVTVD----- 423
         |   :                              ||| :::
Db 686 T----ATVQSNLSWSVDVPAADIQALGNGDLTVNASVTNGVGNTGSGSRDITIDANLPG 740

Qy 424 ---------PVQFNPVQNLTGSAVG                              439
                  ::
Db 741 LRVDTVAGDDVINSIEHNQALVITGSSSG                         769


RESULT 12
B75622
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: B75622
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75622
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-691 <WHI>
A;Cross-references: GB:AE001826; NID:g6460827; PIDN:AAF12628.1; PID:g6460924; TIGR:DRB0
C;Genetics:
A;Gene: DRB0037
A;Map position: megaplasmid
A;Genome: plasmid
```

```
A;Note: plasmid MP1

Query Match       5.1%;   Score 124.5;  DB 2;  Length 691;
Best Local Similarity   22.1%;   Pred. No. 0.67;
Matches 110;  Conservative  45;  Mismatches 185;  Indels 157;  Gaps 23;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWT------TTPPPGGTSFAGHNSAICA  54
         :::  || |||                         |   ||| |   | : :
Db 231 PNPLPPTGTTPGTTPGTGTG-PVGSGNGPAESTGPAYTGIGSTSGTQPPVVVDASGNQYAYPK 289

Qy  55 SSASYINFEGP------QNPD---NYLVTPELS-LPNGGTLTFWVCAQDAN-------Y  96
         :::       |   ||     | | :||   :     :|||
Db 290 ADAETVNPDSVTMTTITNPTAATYELVPDLSGLPAGVTVTF----TDANGNPLPDTD   345

Qy  97 ASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGT 156
         |  :                  ||                           :
Db 346 GNGRPEVTAGPNGGTA------------TYRVVVTYPDTESAAAVAG---PIRIPVGV   388

Qy 157 K--------YVAFRHFGCTDFFWINLDDVEIKANGKRADFT--------ETFES      194
                  :|                     ||||        :
Db 389 DGNRDGIVDATVTYNVLLSNLKFGNTNGTALGVSDVPVTRTVTQPSVTTAVVFPMDLFND 448

Qy 195 STH--------GEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWN    242
         |||        | |                             :
Db 449 GAYDGNYALSGSTPIGPVXYATNPDTDGDG----VLSPAELAALPAE----IAS---   495

Qy 243 GMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNG 302
                      :||                    ||
Db 496 -------------YAVVTIPAGQAPGDYMVT--QTATG                     530

Qy 303 INKGGARFGLSTEADGAKPQ-SVWIERTVDLPAGT--------KYVAFRHYN        345
              |           |                          ||
Db 531 SLSGTTKSFNTDKVTVTSPNGSLLIAKRVTTPGTTTPSLNATANPGDAVSYTVTATNNYN 590

Qy 346 CSDLNYILLDDIQFTMG-------GSPTPTDYTYTV----YRD---GTKIKEGL     385
         :                        |:::                ||
Db 591 TSLYGLVLRDPSSNNLGSFSSSNVFGFIKPSSLRATVSGVSGATVLYRTSNLNTWAAQPT 650

Qy 386 TB--TTFEEDGVATGNH                                          400
         |    ||
Db 651 VDANTTWVEVGVDTNNN                                          667


RESULT 13
B97033
uncharacterized protein, related to enterotoxins of other Clostridiales [imported] - Cl
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: B97033
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: B97033
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2817 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK79053.1; PID:g15023993; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1079

Query Match       5.1%;   Score 124.5;  DB 2;  Length 2817;
Best Local Similarity   22.2%;   Pred. No. 4.2;
Matches 100;  Conservative  41;  Mismatches 149;  Indels 161;  Gaps 24;

Qy   3 PNPGTTTLSESFENGIPASWKTIDADGDG--NNWTTTP------PPGGTSFAG------  47
         ||      : | :                                  :: |||
Db 926 PNYGYMLLGYNYING----KYYYFDNDGVIQTGWVTDRSSKYYLDPSGAAVTGFQNING  980

Qy  48 -----HNSAICASSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYA 102
              ::|    :
```

```
RESULT 14
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B85547
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85547
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5188 <STO>
A;Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:Z06
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0615

Query Match             5.1%;   Score 124;   DB 2;   Length 5188;
Best Local Similarity  20.7%;   Pred No. 10;
Matches 123;   Conservative 67;   Mismatches 181;   Indels 222;   Gaps 34;

Qy     6  GTT------TLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS--- 56
              |||      ||:||    ||:|    |||    |||  ||||
Db   965  GTTTAQAGQTLTVTLNNNYQT--TVLADG--TWSVNVP------AADLSGLTASSYTV 1013

Qy    57  -ASYINFEG-PQNPDNYLIV--TPELSLPNGGTLTFWVCAQDANYASEH-YAVYASSTG 109
Db  1014  TATVSDKAGNPASADHALVVDITAPDLTINT------VAGDDIINAIEHGQALVVSGTS 1066

Qy   110  NDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDF 169
Db  1067  TGRA--AGDVVTVLNGKNYTTIIDA------SGNW--SVGIPAA------ 1101

Qy   170  FWINLDDVEIKANGKR---------ADFTETEFSSTHGEAPAEWTTIDADG------DGQG 215
Db  1102  --DVTALATGSQTITASLSDRAGNSDSTTHD-----VTVDLSGPLTINTVSGDD 1149

Qy   216  WLCLS------SGQLDWLTAHGGTNVV------------ASFSWNG------MALNP 248
Db  1150  IINAAEIVAQTISGQVTGTAVAGNTVIVTIGGNQYNATVQSDLSWSVSVPANVLQALGN 1209

Qy   249  DNYLISKDVTGATK---VKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV------V 295

Db   981  DKYYFNSSGIMQTGLVYV------NPDYYGFD-----DNGHILT------GMHSINGYI 1022
Qy   103  VYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQ---LPAGTKY 158
Db  1023  YYFDSTGKAQKGFV-----------TYLGKTYFNTNMYTGFVNANNNL 1060
Qy   159  VAFRHFGCTDFFWINLDDVEIKANGKRADFTETEFSSTHGEAPAEWTTIDADGDGQGWLC 218
Db  1061  YYFDNEGVMQTGWINY-----NSNRYYFSAT------GASVTGFQTI---DGNKYCF 1103
Qy   219  LSSGQL--DWLTAHGGTNVVASFSWNGMALNP-------DNYLISKDVTGAT- 261
Db  1104  DSNGAIYTDVVTINGST-----YGFNTDGIMLTGWQTIRYNRGYSSYFNTYFNSDGTAKTG 1160
Qy   262  ----KVKYYYAVNDG-------FPGDHYAVMISKTGTNAGDFTVVFEETPNGINK- 305
Db  1161  FFTYLNKTYYFNPSDGRMLQGYQYINGNHYYF------APDGTMQTGWITNGSSKYYL 1212
Qy   306  --GGARFGLST-----------EADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 354
Db  1213  DPSGAAVTGLQTINGNKYCFDSNGLLQHN------GIFYIGNTYYG-SDNNGIML 1260
Qy   355  DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGL 385
Db  1261  TGLQLING------YLYCFNSDGS-VKTGL 1283
```

```
Db  1210  GELTISASLTNSANNTGTATHDIVIDANLPG-------LRVDTVAGDDVINSIEHTQALV 1262
Qy   296  FEETPNGINKGA----------RFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCS 347
Db  1263  ITGSSSGLAAGAALTVINSVTYGATVLADGSWSVGVPVADVTNWPAGTVNIAVSGTNTA 1322
Qy   348  --------------------------DLNYILLDDI--------QFTMGGSPTPTD- 369
Db  1323  GTTTSISHPVTVDLAAVAIIINTLSTDDVINAAEKGSDLQLSGTTSGVEAGQTITVIFGG 1382
Qy   370  --YTVTVYRDGTKIKEGLT-------ETTFEEDGVA------TGN------HEYCVEVKYT 409
Db  1383  KSYTTTVAADNT--WGLTIPAVDVATLPDGAANVQASVSNVAGNSTQATHAYSVD--- 1435
Qy   410  AGVSPKECVNTVDPVQFNPVQNL-TGSAV-----------GQKVTLKWDAPN 450
Db  1436  -ATAP----SVTINTIATDDILNAABAGSALTISGTSTAEAGQTVTVTLNGVN 1483
```

```
RESULT 15
T13256
tail-host specificity protein homolog - Lactococcus lactis phage BK5-T
C;Species: Lactococcus lactis phage BK5-T
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C;Accession: T13256
R;Boyce, J.D.; Davidson, B.E.; Hillier, A.J.
Appl. Environ. Microbiol. 61, 4089-4098, 1995
A;Title: Sequence analysis of the temperate Lactococcus lactis bacteriophage BK5-T and
A;Reference number: Z17646; MUID:96064422; PMID:8526523
A;Accession: T13256
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1904 <BOY>
A;Cross-references: EMBL:L44593; NID:g928826; PID:g928828; PIDN:AAA98579.1

Query Match             5.1%;   Score 123.5;   DB 2;   Length 1904;
Best Local Similarity  22.4%;   Pred. No. 3;
Matches 105;   Conservative 50;   Mismatches 144;   Indels 169;   Gaps 30;

Qy     2  NPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPP----PGGTSFAG------H 48
Db   493  NVNDGADGKTIPYFH----TAW-SYSADGT-DRFTTTYPNLNLLEGSKDFSGGWLIVERSE 546
Qy    49  NSAICASSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASST 108
Db   547  NDGI-YKGLTVKKSEGPG-----ITKQFIAPKEGIYTF----------SAVVKSS 585
Qy   109  GNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQ--LPAGTK-YVAFRHFG 165
Db   586  GNNA-NIIRAVTLNNVTGKIV---PDKSMG--INFGWLRDSFQVTLKAGDKIYAQYNVAG 639
Qy   166  ----------------CTDF---------------------------WINLDDVE 178
Db   640  SGVLWNAGHKWEEGSAATDLTSNYPNLNLLEGSKKYTKDNPRIQSSANDGWTTVDDVF 699
Qy   179  IKANGKRADFTETEFSSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVAS 238
Db   700  VK-NLKAGTYTMS--------GKADAPWTNHDTSGANKGKVGL------WLVS------ 737
Qy   239  FSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV---- 294
Db   738  --TTPGLGVN----IS---LGETVPKTIEVPKDG----DYCVRVN-TYSNGKDIEAHKFW 783
Qy   295  VFEETPNGINKGGARFGLSTEADGAKPQSVWIE-----RTVDLPAGT-KYVAFRHYNCSD 348
Db   784  DFKLEPGSI---ATPWMPSKNELEQLIPWMPSLSEVTTADYPSYIGQYTDFIQYD--- 836
Qy   349  LNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVA 396
Db   837  ------SAKPSDYTWSLIR-GNDGKDGAT----GKDGVA 864
```

Search completed: May 18, 2004, 11:47:47

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 18, 2004, 11:35:14 ; Search time 8.22454 Seconds
                              (without alignments)
                              2848.981 Million cell updates/sec

Title:          US-08-570-311-16
Perfect score:  2443
Sequence:       1 PNPNPGTTLSESFENGIPA..........QNLTGSAVGQKVTLKWDAPN 450

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:    141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SwissProt_42.*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2443	100.0	2628	1	HGA2_PORGI	Q51845 porphyromon
2	2434	99.6	2164	1	HGA1_PORGI	P59915 porphyromon
3	1460	59.8	989	1	PRTH_PORGI	P46071 porphyromon
4	266	10.9	991	1	CPG1_PORGI	P28784 porphyromon
5	126	5.2	721	1	OGP_MOUSE	Q62010 mus musculu
6	121.5	5.0	465	1	SLAP_LACBR	Q05044 lactobacilli
7	121	5.0	1090	1	GUXB_CELFI	P50899 cellulomona
8	120	4.9	2660	1	YEEJ_ECO57	Q8X8V7 escherichia
9	117.5	4.8	1256	1	ATL_STAAU	P52081 staphylococ
10	116.5	4.8	331	1	PME_ASPAC	Q12535 aspergillus
11	116.5	4.8	1385	1	C5AA_BACUD	Q45760 bacillus th
12	114.5	4.7	607	1	YK28_ARCFU	O28251 archaeoglob
13	112.5	4.6	987	1	YD94_METJA	Q58789 methanococc
14	111.5	4.6	1122	1	ADP1_MYCGA	Q49379 mycoplasma
15	110.5	4.5	1034	1	BGAL_BACME	O52847 bacillus me
16	109.5	4.5	282	1	PRTA_ASPNG	P24665 aspergillus
17	109.5	4.5	1953	1	BIGA_SALTY	P25927 salmonella
18	109	4.5	872	1	GUXA_CELFI	P50401 cellulomona
19	108.5	4.4	1260	1	ALS1_CANAL	P46590 candida alb
20	106.5	4.4	1200	1	HYAL_STRPU	O76536 strongyloce
21	106	4.3	827	1	XANP_XANS2	Q60106 xanthomonas
22	106	4.3	2812	1	ZAN_HUMAN	Q9Y493 homo sapien
23	105.5	4.3	524	1	CHID_BACCI	P27050 bacillus ci
24	105.5	4.3	1534	1	YFAS_ECO57	Q8XE35 escherichia
25	105	4.3	1045	1	PRTS_SERMA	P09489 serratia ma
26	105	4.3	1300	1	120K_RICRI	P14914 rickettsia
27	105	4.3	1654	1	OMPB_RICRI	Q53047 r outer mem
28	105	4.3	2249	1	OMPA_RICRI	P15921 rickettsia
29	104.5	4.3	569	1	YAP3_YEAST	P32329 saccharomyc
30	104.5	4.3	1122	1	ADP2_MYCGA	Q9rem8 mycoplasma
31	104.5	4.3	2124	1	PGCA_RAT	P07897 rattus norv
32	104	4.3	609	1	HAPT_VIBCH	P24153 vibrio chol
33	104	4.3	1045	1	GUNB_CELFI	P26225 cellulomona
34	103.5	4.2	1861	1	APU_THETU	P38536 t amylopull
35	103	4.2	435	1	AM3D_ORYSA	P27933 oryza sativ
36	103	4.2	548	1	THER_BACST	P06874 bacillus st
37	103	4.2	1044	1	ITA8_CHICK	P26009 gallus gall
38	103	4.2	1157	1	C8AA_BACUK	Q45704 bacillus th
39	103	4.2	2725	1	FLNC_HUMAN	Q14315 homo sapien
40	102.5	4.2	551	1	AMYB_THETU	P19984 thermoanaer
41	102.5	4.2	1220	1	C5AC_BACTU	P56955 bacillus th
42	102.5	4.2	2132	1	PGCA_MOUSE	Q61282 mus musculu
43	102	4.2	1289	1	C5AB_BACUD	Q45753 bacillus th
44	101.5	4.2	1045	1	PRTT_SERMA	P29805 serratia ma
45	101.5	4.2	1341	1	VG37_BPT2	P07067 bacteriopha

ALIGNMENTS

```
RESULT 1
HGA2_PORGI
ID   HGA2_PORGI   STANDARD;   PRT;   2628 AA.
AC   Q51845;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97047672; PubMed=8926061;
RA   Han N., Whitlock J. V. Progulske-Fox A.;
RT   "The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381
RT   contains four large, contiguous, direct repeats.";
RL   Infect. Immun. 64:4000-4007(1996).
CC   -!- FUNCTION: Agglutinates erythrocytes.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U41807; AAB17128.1; -.
DR   PIR; T28651; T28651.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   2628       HEMAGGLUTININ A.
FT   DOMAIN       25    539       PEPTIDASE C25-LIKE 1.
FT   DOMAIN      540    995       PEPTIDASE C25-LIKE 2.
FT   DOMAIN      996   1451       PEPTIDASE C25-LIKE 3.
FT   DOMAIN     1452   1907       PEPTIDASE C25-LIKE 4.
FT   DOMAIN     2074   2628       PEPTIDASE C25-LIKE 5.
SQ   SEQUENCE   2628 AA;   283324 MW;   61C4DB32540C99DA CRC64;

Query Match             100.0%;   Score 2443;   DB 1;  Length 2628;
Best Local Similarity   100.0%;   Pred. No. 1.1e-163;
Matches   450;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSPAGHNSAICASSASYI 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    500 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSPAGHNSAICASSASYI 559

Qy     61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 619
```

```
Qy  121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  620 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 679

Qy  181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  680 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 739

Qy  241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGPPGDHYAVMISKTGTNAGDFTVVFEETP 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  740 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGPPGDHYAVMISKTGTNAGDFTVVFEETP 799

Qy  301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  800 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 859

Qy  361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  860 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 919

Qy  421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
        ||||||||||||||||||||||||||||||
Db  920 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 949
```

RESULT 2
HGA1_PORGI

```
ID  HGA1_PORGI   STANDARD;      PRT;  2164 AA.
AC  P59915;
DT  10-OCT-2003 (Rel. 42, Created)
DT  10-OCT-2003 (Rel. 42, Last sequence update)
DT  15-MAR-2004 (Rel. 43, Last annotation update)
DE  Hemagglutinin A precursor.
GN  HAGA OR PG1837.
OS  Porphyromonas gingivalis (Bacteroides gingivalis).
OC  Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC  Porphyromonadaceae; Porphyromonas.
OX  NCBI_TaxID=837;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=W83;
RX  MEDLINE=22829867; PubMed=12949112;
RA  Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA  Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
RA  Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
RA  Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA  Dewhirst F.E., Fraser C.M.;
RT  "Complete genome sequence of the oral pathogenic bacterium
RT  Porphyromonas gingivalis strain W83.";
RL  J. Bacteriol. 185:5591-5601(2003).
CC  -!- FUNCTION: Agglutinates erythrocytes (By similarity).
CC  -!- SIMILARITY: Belongs to peptidase family C25.
DR  EMBL; AE017178; AAQ66831.1; ALT_INIT.
DR  TIGR; PG1837; -.
KW  Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
KW  Complete proteome.
FT  SIGNAL     1    25       POTENTIAL.
FT  CHAIN     26  2164       HEMAGGLUTININ A.
FT  DOMAIN    26   539       PEPTIDASE C25-LIKE 1.
FT  DOMAIN   540   991       PEPTIDASE C25-LIKE 2.
FT  DOMAIN   992  1443       PEPTIDASE C25-LIKE 3.
SQ  SEQUENCE 2164 AA;  233387 MW;  6DFAB2282586C63 CRC64;
```

Query Match   99.6%; Score 2434; DB 1; Length 2164;

Best Local Similarity 99.6%; Pred. No 3.7e-163;
Matches 448; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy    1 PNPNPGTTLSESFPENGIPASWKTIDADGDGNWTTTPPPGGTSFAGHNSAICASSASYI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db  952 PNPNPGTTLSESFPENGIPASWKTIDADGDGNWTTTPPPGGTSFAGHNSAICVSSASYI 1011

Qy   61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDASNFANALL 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1012 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDASNFANALL 1071

Qy  121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1072 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 1131

Qy  181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1132 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 1191

Qy  241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGPPGDHYAVMISKTGTNAGDFTVVFEETP 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1192 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGPPGDHYAVMISKTGTNAGDFTVVFEETP 1251

Qy  301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
        |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db 1252 NGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 1311

Qy  361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1312 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 1371

Qy  421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
        ||||||||||||||||||||||||||||||
Db 1372 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 1401
```

RESULT 3
PRTH_PORGI

```
ID  PRTH_PORGI   STANDARD;      PRT;   989 AA.
AC  P46071;
DT  01-NOV-1995 (Rel. 32, Created)
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Protease prtH (EC 3.4.22.-).
GN  PRTH.
OS  Porphyromonas gingivalis (Bacteroides gingivalis).
OC  Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC  Porphyromonadaceae; Porphyromonas.
OX  NCBI_TaxID=837;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=W83;
RX  MEDLINE=95012612; PubMed=7927685;
RA  Fletcher H.M., Schenkein H.A., Macrina F.L.;
RT  "Cloning and characterization of a new protease gene (prtH) from
RT  Porphyromonas gingivalis.";
RL  Infect. Immun. 62:4279-4286(1994).
RN  [2]
RP  ERRATUM.
RA  Fletcher H.M., Schenkein H.A., Macrina F.L.;
RL  Infect. Immun. 62:5707-5707(1994).
CC  -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
CC      P.GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
CC      IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
CC      AND IS A VIRULENCE FACTOR.
CC  -!- SUBCELLULAR LOCATION: In membrane vesicles.
CC  -!- SIMILARITY: Belongs to peptidase family C25.
```

RL   J. Biol. Chem. 267:18896-18901(1992).
CC   -!- FUNCTION: Thiol protease which is believed to participate in
CC   intracellular degradation and turnover of proteins. Its
CC   proteolytic activity is a major factor in both periodontal tissue
CC   destruction and in bacterial host defense mechanisms. Activates
CC   complement C3 and C5.
CC   -!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
CC   immunoglobulins, with a preference for Arg in P1, and hydrophobic
CC   residues in P2 and P3.
CC   -!- ENZYME REGULATION: Requires cysteine for activation and Ca(2+)
CC   and/or Mg(2+) for stabilization. It is stimulated by glycine-
CC   containing dipeptides. It is resistant to inhibition by proteinase
CC   inhibitors in human plasma.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; D26470; BAA05484.1; -.
DR   PIR; 140229; 140229.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
KW   Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
FT   SIGNAL   1   24
FT   PROPEP   25   227
FT   CHAIN   228   991   GINGIPAIN R1.
FT   ACT_SITE   438   438   PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   471   471   NUCLEOPHILE (BY SIMILARITY).
FT   CONFLICT   264   265   RT -> TK (IN REF. 2).
SQ   SEQUENCE   991 AA; 108782 MW; 03EE3F43CEBE2544 CRC64;

Query Match
Best Local Similarity 10.9%; Score 266; DB 1; Length 991;
Matches 114; Conservative 50; Mismatches 164; Indels 112; Gaps 19;

Qy   24 TIDADGDGNNWTTTPPPGGTSPAGHNSAICASSASYINFEGPQNPDNYLVTPELSLPNGG 83
          ::   :
Db 595 SVNVSCDYNGAIATISANGKMF---GSAVVENGTATINLTG----------LTNES 637

Qy   84 TLTFWVCAQDANYASEHYAVYASSTG--NDASNFANALLEEVLTAKTVVTAPEAIRGTRV 141
          ::   :|| :
Db 638 TLTLTV---VGYNKETVIKTINTNGEPNPYQPVSN-----LTATT------QQGKV 679

Qy 142 QGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV-EIKANGKRADFTETFESSTHGEA 200
          :|| :
Db 680 TLKWDAPSTKTNATTN-TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWNDGS- 737

Qy 201 PAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWN----GMALNPDNYLISK 255
           : :   :
Db 738 -GYQILLDADHDQYGQVIPSDTHTLMPNCSVPANLFAPFEYTVPENADPSCCSPTNMIM-- 794

Qy 256 DVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTE 315
          | :
Db 795 DGTASVNI---------PAGTY-----------DFAI----------------------- 811

Qy 316 ADGAKPQS--VWI-----ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP 365
                                  :
Db 812 ---AAPQANAKIWIAGQGPTKEDDVVFEAGKKY----HFLMKKMGSGDGTELTISEGGG- 863

Qy 366 TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKCCVNVTVDPV 425
          ||||||||||||||||||||||| : |||:||||||||| ||||| || ||||
Db 864 --SDYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHYCVEVKYAAGVSPKVCVDYIPDGV 921

CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; L27483; AAA51298.1; -.
DR   HSSP; P23882; 1FMT.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR002376; formyl_transf.
DR   Pfam; PF00551; formyl_transf; 1.
KW   Hydrolase; Thiol protease; Repeat; Virulence.
FT   REPEAT   270   323
FT   REPEAT   528   581
SQ   SEQUENCE   989 AA; 110238 MW; FA85FE83AC8944C CRC64;

Query Match
Best Local Similarity 85.2%; Score 1460; DB 1; Length 989; Pred. No. 4.5e-95;
Matches 277; Conservative 11; Mismatches 29; Indels 8; Gaps 3;

Qy 130 VTAPEAIRGTRVQGTWYQKTVDLPAGTKYVAFRHFG----CTDFFWINLDDVEIKANGKR 185
          :||||||| :|:||| |||||||||||||
Db   34 VRSPEAIRG-RIQGTWRQKTVDLPAGTEICCFPSLPKAPICSTSTLMRL--RSKTNAKR 89

Qy 186 ADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMA 245
          |||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db   90 ADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMA 149

Qy 246 LNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINK 305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150 LNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINK 209

Qy 306 GGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP 365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 210 GGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP 269

Qy 366 TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPV 425
          ||||||||||||||||||||||||||||||||||||||||||||||||:|||||:
Db 270 TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTVNST 329

Qy 426 QFNPVQNLTGSAVGQKTLKWDAPN 450
          ||||: || || ||| :
Db 330 QFNPVKNLKAQPDGGDVVLKWEAPS 354

RESULT 4
CPG1_PORGI
ID   CPG1_PORGI   STANDARD;   PRT;   991 AA.
AC   P28784; Q45168;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
DE   (RGP-1).
GN   RGPA OR RGP1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 228-290 AND 517-541.
RC   STRAIN=381;
RX   MEDLINE=95168884; PubMed=7864651;
RA   Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
RA   Ikehara Y.;
RT   "Structural characterization of arginipain, a novel
RT   arginine-specific cysteine proteinase as a major periodontal
RT   pathogenic factor from Porphyromonas gingivalis.";
RL   Arch. Biochem. Biophys. 316:917-925(1995).
RN   [2]
RP   SEQUENCE OF 228-270.
RC   STRAIN=HG66;
RX   MEDLINE=92406812; PubMed=1527017;
RA   Chen Z., Potempa J., Polanowski A., Wikstrom M., Travis J.;
RT   "Purification and characterization of a 50-kDa cysteine proteinase
RT   (gingipain) from Porphyromonas gingivalis.";

```
Qy    426 QFNPVQN-LTGSAVGQKVTL 444
          |||: :||   |:.:||
Db    922 ADVTAQKPYTLTVVGKTITV 941
```

RESULT 5
OGP_MOUSE
ID   OGP_MOUSE   STANDARD;   PRT;   721 AA.
AC   Q62010;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Oviduct-specific glycoprotein precursor (Oviductal glycoprotein)
DE   (Oviductin) (Estrogen-dependent oviduct protein).
GN   OVGP1 OR OGP OR CHIT5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR; TISSUE=Oviduct;
RX   MEDLINE=96115001; PubMed=7492680;
RA   Sendai Y., Komiya H., Suzuki K., Onuma T., Kikuchi M., Hoshi H.,
RA   Araki Y.;
RT   "Molecular cloning and characterization of a mouse oviduct-specific
RT   glycoprotein.";
RL   Biol. Reprod. 53:285-294(1995).
CC   -!- FUNCTION: Binds to oocyte zona pellucida in vivo. May play a role
CC   in the fertilization process and/or early embryonic development.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- TISSUE SPECIFICITY: Epithelial cells of the oviduct.
CC   -!- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; D32137; BAA06863.1; -.
DR   MGD; MGI:106661; Ovgp1.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR001579; Glyco_hydro_18AS.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   ProDom; PD000471; Glyco_hydro_18; 1.
DR   SMART; SM00636; Glyco_18; 1.
DR   PROSITE; PS01095; CHITINASE_18; FALSE_NEG.
KW   Glycoprotein; Fertilization; Repeat; Signal.
FT   SIGNAL   1   21   POTENTIAL.
FT   CHAIN   22   721   OVIDUCT-SPECIFIC GLYCOPROTEIN.
FT   DOMAIN   486   632   21 X 7 AA TANDEM REPEATS OF S-K-T-T-
FT   [TAP]-G-[IV].
FT   CARBOHYD   402   402   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   442   442   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   469   469   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   721 AA; 78807 MW; 37246C8F01665652 CRC64;

Query Match   5.2%;   Score 126; DB 1; Length 721;
Best Local Similarity 18.6%;   Pred. No. 0.22;
Matches 103;   Conservative 53;   Mismatches 178;   Indels 220;   Gaps 22;

Qy   45 FAGHNSAICA------SSASYINF---------------------- 62
            ||||| | :: |:
Db   222 FTGHNSPLFSLPEDSKSSAYAMNYWRKLGTPADKLIMGFPPTYGRNFYLLKESKNGLQTAS 281

Qy   63 EGPQNPDNYLIVTPELSLPNGGTLTFWVCAQDANYASEHYAVTASSTGNDASNFANALLEE 122
            :|| ||:| .:| ||: |:.:|:.:
Db   282 MGPASPGKY-------TKQAGFLAYYSWVCSFVQRAKKHWIDYQ--------------- 317

Qy   123 VLTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV------ 177
```

```
Qy    318 --------------YVPYAFKGKEWLG--YDDTISFSYKAMYVKREHFGGAMVWTLDMDDVRGTFC 367
                             |||  .|       | .  :|:: :|||
Db    318 ------------------------------EIKANGKRAD----FTETEFSST------------ 196
```

```
Qy    178 -------------------------------------EIKANGKRAD----FTETEFSST------------ 196
```

```
Db    368 GNGPPPLVIHILNELLVQTESNSTPLPQFWTSSVNASGPQSENTALTEVLTTDTIKILPP 427
              :||  | ||| |:|  ||:   :   |:.:                    -NVV 236
Qy    197 HGEAPA----------------------EWTTIDADGDGQWLCLSSGQLDWLTAHGGT 236
```

```
Db    428 GGEAMTTEVRRRYENMTTVPSDGS---------------------VTPGGTASPRKGHAVTPENNTM 472
             ||  :                    : .||
Qy    237 AS----------------------FSWNGMALNPDNYLISKDVTGATKV----KYYYAVNDGFPGDHYAVM 281
```

```
Db    473 AAEAKTMSTLDFFSKTTTGVSKTTTGISKTTTGVSKATAGISKTIPEISKATA 532
             |||| | :||  :|||  :|:|||  :||   |:|||  :|||  :||
Qy    282 -ISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEADGAKPQSVWIER-TVDLPAGTKYV 339
```

```
Db    533 GVSKTTTGVSKTTTGISKTITGVSK--TTTGISKTTTGISKTTTGISKTTTGVSKTTTGI 590
             :|||  :||   :||   |:|||   :|||   :|||   :|||   :|||   ||
Qy    340 AFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTVYVYRDGTKIKEGLTETT---------- 389
```

```
Db    591 SKTTTGIS--------------QTTTGISKTTTDISKTT-TGISKTTTPGISKTTPGMTVIVQTQ 639
             :|||||               |:||  :|:|||  :|||  |:||||| :|||  |  :   ||
Qy    390 -FEEDGVATGNHEYCVEVK----------YTAGVSPKE------------CVNVTVDPVQFNPVQN 432
```

```
Db    640 ANEAETTATMDHQSVTPTEMDTTLFYLKMTPSEKETSRKKTMVLEKAFVSPREMSATPN 699
             | .|| | || :|  | .|  | |  :: :|  |||  :||||
Qy    433 LTGSAVGQKVTLKW 446
             |||||
Db    700 ------GQSKTLKW 707
```

RESULT 6
SLAP_LACBR
ID   SLAP_LACBR   STANDARD;   PRT;   465 AA.
AC   Q05044;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   S-layer protein precursor (Surface layer protein).
OS   Lactobacillus brevis.
OC   Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=1580;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=ATCC 8287 / DSM 20556;
RX   MEDLINE=93054358; PubMed=1429463;
RA   Vidgren G., Palva I., Pakkanen R., Lounatmaa K., Palva A.;
RT   "S-layer protein gene of Lactobacillus brevis: cloning by polymerase
RT   chain reaction and determination of the nucleotide sequence.";
RL   J. Bacteriol. 174:7419-7427(1992).
CC   -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC   of proteins which coat the surface of bacteria.
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a
CC   S-layer with tetragonal symmetry.
CC   -!- SIMILARITY: SOME, TO THE S-LAYER PROTEIN OF L.ACIDOPHILUS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Z14250; CAA78618.1; -.
DR   PIR; A47023; A47023.
KW   Signal; Glycoprotein; Cell wall; S-layer.
FT   SIGNAL   1   30
FT   CHAIN   31   465   S-LAYER PROTEIN.
SQ   SEQUENCE   465 AA; 48159 MW; 2BE240339265A2C CRC64;

```
Query Match        5.0%;  Score 121.5;  DB 1;  Length 465;
Best Local Similarity  19.8%;  Pred. No. 0.25;
Matches  99;  Conservative  48;  Mismatches 201;  Indels 153;  Gaps  20;

Qy    7 TTTLSESFEN-GIPASWKTIDA----DGDGNWWTTPPGGTSFAGHNSAICASSASYIN 61
Db   26 TTASAKSYATAGAYSTLKTDAATRNVEATGTNALYTKP--GTV---KGAKVVASKATMAK 80

Qy   62 FEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLE 121
Db   81 LASSKKSADYFRAYGVKTTNRGSVYYRVVTMDGKYRGVYG---GKSDTAFAGGi-- 132

Qy  122 EVLTAKIVVTAPEAIRGTRVQGTWYQKTVQ-----LPAGTKYVAFRHFGCTDFFWINLDD 176
Db  133 -----KSAETTTKADMPARTTGFYLTDTSKNTIWTAPKYTQYKASK-------------- 173

Qy  177 VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDG-QGWLCLSSGQLDWLTAHGGTNV 235
Db  174 VSLYGVAKDTKFPTVDQAATKTREGSLYHVTATNGSGISGWIYAGKG---FSTTATGTQV 230

Qy  236 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 295
Db  231 LGGLSTDK---------------SVTATNDNSVKIVRTTD------------GTQVGSNTWV 266

Qy  296 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYV-------AFRHYNCS 347
Db  267 ------------TSTDGTKAGSKVSDKAADQTALEAYINANKPSGYTVTNPNAA 308

Qy  348 DLNY------------ILLDDIQFTMGGSPTPTDYT--------------YTVY 375
Db  309 DATYGNTVYATVSQAATSKVALKVSGTPVTTALTTADANDKVAANDTTANGSSVAGSTVY 368

Qy  376 RDGTKIKEGLIETTFE-----------------EDGVATGNHEYCVEV-----KYTAGVSPK 415
Db  369 AAGTKLAQLITDLTGEKGQVVTLTAIDTDLEDATFTGTTTYYSDLGKAYHYTYTYNKDSA 428

Qy  416 ECVNVTVDPVQFNPVQNLTGS 436
Db  429 ASSNAS---TQFG--SNVTGT 444
```

RESULT 7
```
ID     GUXB_CELFI   STANDARD;   PRT;   1090 AA.
AC     P50899;
DT     01-OCT-1996  (Rel. 34, Created)
DT     01-OCT-1996  (Rel. 34, Last sequence update)
DT     10-OCT-2003  (Rel. 42, Last annotation update)
DE     Exoglucanase B precursor (EC 3.2.1.91) (Exocellobiohydrolase B)
DE     (1,4-beta-cellobiohydrolase B) (CBP120).
GN     CBHB OR CENE.
OS     Cellulomonas fimi.
OC     Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC     Micrococcineae; Cellulomonadaceae; Cellulomonas.
OX     NCBI_TaxID=1708;
RN     [1]
RP     SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
RC     STRAIN=ATCC 484;
RX     MEDLINE=96003898; PubMed=7575482;
RA     Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT     "Cellobiohydrolase B, a second exo-cellobiohydrolase from the
RT     cellulolytic bacterium Cellulomonas fimi.";
RL     Biochem. J. 311:67-74(1995).
RN     [2]
RP     SEQUENCE OF 54-75.
RX     MEDLINE=93209933; PubMed=8458833;
RA     Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT     "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RT     D (CenD), a family A beta-1,4-glucanase.";
RL     J. Bacteriol. 175:1910-1918(1993).
RN     [3]
RP     SEQUENCE OF 54-78.
RX     MEDLINE=94197708; PubMed=8147863;
RA     Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.,
RA     Warren R.A.J., Miller R.C. Jr.;
RT     "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi
RL     CenB, a member of a new family of beta-1,4-glucanases.";
CC     Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
CC  -!- FUNCTION: Hydrolyze cellohexaose to a mixture of cellotetraose,
CC     cellotriose and cellobiose, with only a trace of glucose. It
CC     hydrolyzed cellopentaose to cellotriose and cellobiose, and
CC     cellotriose to cellobiose, but it did not hydrolyze cellotriose.
CC     Has also weak endoglucanase activity. Hydrolyzes glucosidic bonds
CC     with inversion of anomeric configuration.
CC  -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC     in cellulose and cellotetraose, releasing cellobiose from the non-
CC     reducing ends of the chains.
CC  -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC  -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC     domain.
CC  -!- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL
CC     HYDROLASES).
CC
CC     This SWISS-PROT entry is copyright. It is produced through a collaboration
CC     between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC     the European Bioinformatics Institute. There are no restrictions on its
CC     use by non-profit institutions as long as its content is in no way
CC     modified and this statement is not removed. Usage by and for commercial
CC     entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC     or send an email to license@isb-sib.ch).
DR     EMBL; L38827; AAB00822.1; -.
DR     PIR; S59077; S59077.
DR     HSSP; P07986; 1EXG.
DR     InterPro; IPR001919; Bac_celose-bind.
DR     InterPro; IPR008965; Cellul_bind.
DR     InterPro; IPR008857; FN_III-like.
DR     InterPro; IPR003961; FN_III.
DR     InterPro; IPR000556; Glyco_hydro_48.
DR     InterPro; IPR008928; Glyco_trans_6hp.
DR     Pfam; PF00553; CBM_2; 1.
DR     Pfam; PF00041; fn3; 3.
DR     Pfam; PF02011; Glyco_hydro_48; 1.
DR     PRINTS; PR00844; GLHYDRLASE48.
DR     ProDom; PD011903; Glyco_hydro_48; 1.
DR     SMART; SM00637; CBD_II; 1.
DR     SMART; SM00060; FN3; 3.
DR     PROSITE; PS00561; CBD_BACTERIAL; 1
KW     Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT     SIGNAL      1     33       POTENTIAL.
FT     PROPEP     34     53
FT     CHAIN      54   1090       EXOGLUCANASE B.
FT     DOMAIN     54    699       CATALYTIC (BY SIMILARITY).
FT     DOMAIN    700    785       FIBRONECTIN TYPE-III 1.
FT     DOMAIN    794    884       FIBRONECTIN TYPE-III 2.
FT     DOMAIN    891    978       FIBRONECTIN TYPE-III 3.
FT     DOMAIN    989   1090       CELLULOSE-BINDING (BY SIMILARITY).
FT     ACT_SITE  513    513       NUCLEOPHILE (BY SIMILARITY).
FT     DISULFID  990   1089       BY SIMILARITY.
SQ     SEQUENCE 1090 AA; 114829 MW; 046BB9D956F2F399 CRC64;
```

```
Query Match      5.0%;   Score 121;  DB 1;  Length 1090;
Best Local Similarity  22.3%;  Pred. No. 0.85;
Matches 106;  Conservative  53;  Mismatches 197;  Indels 120;  Gaps  26;

Qy   20 ASWKT---IDADGDGNWWTTTPPGG------------TSFAGHNSAICASSASYINFEGPQNPD 69
Db  515 ASWKVPSELKWTGKPDTWNAAAPTGNPGLITVEVTSY-GQDGVAADTARALLFYAAKSGD 573

Qy   70 NYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTV 129
Db  574 TASRDKAKALLD----AIWANNQDP------LGVSAVETRGDYKRPDDTYVAN----GDG 619

Qy  130 VTAPEAIRGTRVQG-----------------TWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDD 176
Db  620 IYIPSGWTGTMPNGDVIKPGVSFLDIRSFYKD---PNWSKVQTFLDGGAEPQFRYHRFW 676
```

```
Qy    177  VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVV 236
                :  |||   |   |:|  :  |     |||||||
Db    677  AQTAVAGALADYARLFDDGT------TTPDTTAP-----TVPTG---LQAGVVTSTE 719

Qy    237  ASFSWNGMALNPDNVLISKDV-TGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGD--F 292
               :||| ||      | |||||                :  ||||
Db    720  ATISWT--ASTDDIRVTGYDVYRGATKV-----------GTATTTSFTDTGLTASTAYAY 766

Qy    293  TVVFEETPNGINKGGARFGLSTRA---DGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 349
               |||     :||:  :|:    :||:           :|:|
Db    767  TVRAFDAAGNVSAPSAALTVTTKATPSDTTAP---------SVPAITS------SSSTA 810

Qy    350  NYILLDDIQFTMGGSPTPTDYT-------YTVVRDGTKIKEGLTETTFEEDGV-ATGNHE 401
                    :  |:| | |          :|||||||
Db    811  NSV-----TIGWSAS-TDNAGGSGLAGYDVYRGATRVAQ-TTALTFTDTGLTASTAYE 861

Qy    402  YCVEVKYTAG--VSPKECVNVTV-----DPVQFNPVQNLTGSAVGQ-KVTLKWDA 448
                |  |||      |:|||          :||:| |:|| ||| |:  |||:||
Db    862  YTVRARDVAGNVSAPSTAVSVTTKSDTTPDTTAPSVPAGLAAMTVTETSVALTWNA 917


RESULT 8
YEEJ_ECO57
ID    YEEJ_ECO57       STANDARD;       PRT;   2660 AA.
AC    Q8X8V7; Q8X2B9; Q8X2C0;
DT    28-FEB-2003 (Rel. 41, Created)
DT    28-FEB-2003 (Rel. 41, Last sequence update)
DT    28-FEB-2003 (Rel. 41, Last annotation update)
DE    Hypothetical protein yeeJ.
GN    Z3135 OR ECS2775/ECS2776.
OS    Escherichia coli O157:H7.
OC    Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC    Enterobacteriaceae; Escherichia.
OX    NCBI_TaxID=83334;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX    MEDLINE=21074935; PubMed=11206551;
RA    Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA    Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA    Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA    Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA    Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA    Welch R.A., Blattner F.R.;
RL    "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL    Nature 409:529-533 (2001).
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=O157:H7 / RIMD 0509952;
RX    MEDLINE=21156231; PubMed=11258796;
RA    Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA    Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA    Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA    Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT    "Complete genome sequence of enterohemorrhagic Escherichia coli
RT    O157:H7 and genomic comparison with a laboratory strain K-12.";
RL    DNA Res. 8:11-22(2001).
CC    -!- SIMILARITY: Contains 16 Big-1 domains.
CC    -!- SIMILARITY: Belongs to the intimin/invasin family.
CC    -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC        frameshift in position 1315.
CC    -----------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -----------------------------------------------------------------------
DR    EMBL; AE005423; AAG57041.1; -.
DR    EMBL; AP002559; BAB36198.1; ALT_FRAME.
```

```
DR    EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR    InterPro; IPR003344; Big_1.
DR    InterPro; IPR003535; Intimin.
DR    InterPro; IPR008964; Invasin_intimin.
DR    InterPro; IPR000601; PKD.
DR    Pfam; PF02369; Big_1; 16.
DR    PRINTS; PR01359; INTIMIN.
DR    SMART; SM00634; BID_1; 16.
DR    SMART; SM00089; PKD; 8.
KW    Hypothetical protein; Repeat; Complete proteome.
FT    DOMAIN     738    834       BIG-1 1.
FT    DOMAIN     840    929       BIG-1 2.
FT    DOMAIN     931   1033       BIG-1 3.
FT    DOMAIN    1042   1132       BIG-1 4.
FT    DOMAIN    1134   1236       BIG-1 5.
FT    DOMAIN    1245   1335       BIG-1 6.
FT    DOMAIN    1337   1439       BIG-1 7.
FT    DOMAIN    1448   1539       BIG-1 8.
FT    DOMAIN    1548   1652       BIG-1 9.
FT    DOMAIN    1653   1750       BIG-1 10.
FT    DOMAIN    1751   1855       BIG-1 11.
FT    DOMAIN    1856   1957       BIG-1 12.
FT    DOMAIN    1963   2056       BIG-1 13.
FT    DOMAIN    2065   2156       BIG-1 14.
FT    DOMAIN    2157   2252       BIG-1 15.
FT    DOMAIN    2254   2355       BIG-1 16.
SQ    SEQUENCE   2660 AA;  280062 MW;  01EB92A08F5C09D2 CRC64;

Query Match             4.9%;  Score 120;  DB 1;  Length 2660;
Best Local Similarity   22.5%;  Pred. No. 3.2;
Matches  105;  Conservative  58;  Mismatches  198;  Indels 106;  Gaps  24;

Qy      7  TTTLSESFE---NGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYINFE 63
           || :||:||       | ||:|:|| |  ::   :||:
Db   1470  TATVKDQFDNEVNNLPVTFSTA-----SSGLTLTPGESNTNESGLAQATLAGVAF---- 1519

Qy     64  GPQNPDNYLVTPELSLPNGG-------TLTF---WVCAQDANYASEHYAVVASSTGNDASNF 115
              :  ||:  :|                |:  :  ::
Db   1520  GEQ------TVTASLANNGASDNKTVHFIGDTAAKIIELTPVPDSIIAGTPQNSSGS- 1571

Qy    116  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKIVQLPAGTKYVAFRHFGCTDFFWINLD 175
                    |: | :|||  ||:||:||  |     ||:| |:
Db   1572  ------VITATVVDNNGPPVKGVTVNFTSNAATAEMTNGGQAVTNEQGKATVTYTNTRS 1624

Qy    176  DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQ-GWLCLSSGQLDWLTAHGGTN 234
              ||  |:  :||  :|        : |: | :||:|:|       |:   | |:|:||
Db   1625  SIE---SGARPD---TVEASLENGSSTLSTSINVNADASTAHLTLLQALFDTVSAGDTTN 1678

Qy    235  V---VASFSWNG--------MALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISK 284
               |    |||               |:  || :: :|||     :  :
Db   1679  LYIEVKDNVGNGVPQQEVTLSVSP-----SEGVTPSNNAIY---TTNHDGNFYA---SF 1726

Qy    285  TGTNAGDFTVV------FEETPNGI-NKGGARFGLSTEADGAKPQSVWIERTVDLPAG 335
           ||:|| |          | |  :    :|||   |::||  |:|
Db   1727  TATKAGVYQVTATLENGDSMQQTTVTVPNVANAEISLAASKDPV------IANNNDLTTL 1780

Qy    336  TKYVAFRHYNC--SDLNYILLDDIQ-FTMGGSPTPTDYTYTVYRDGTKI-----KEGLTE 387
              | :|||:|    ||:|: ::|:|    ||:|||| ||||||||:|
Db   1781  TATVADTEGNAIANSEVTFTLPEDVRANFTLG---------DGGKVVTDTEGKAK 1826

Qy    388  TTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLVT 434
                   ::  | :  :| :|:  |||| :  ||::| |||:|:
Db   1827  VTLK--GTKAGAH--TVTASMAGGKSEQLVVNFIADTLTAQVNLNVT 1869


RESULT 9
ATL_STAAU
ID    ATL_STAAU        STANDARD;       PRT;   1256 AA.
AC    P52081;
DT    01-OCT-1996 (Rel. 34, Created)
DT    01-OCT-1996 (Rel. 34, Last sequence update)
DT    28-FEB-2003 (Rel. 41, Last annotation update)
DE    Bifunctional autolysin precursor [Includes: N-acetylmuramoyl-L-alanine
```

```
DE    amidase (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
DE    acetylglucosaminidase (EC 3.2.1.96)].
GN    ATL.
OS    Staphylococcus aureus.
OC    Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX    NCBI_TaxID=1280;
RN    [1]
RP    SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
RC    STRAIN=RN450;
RX    MEDLINE=95116542; PubMed=7816834;
RA    Oshida T., Sugai M., Komatsuzawa H., Hong Y.-M., Suginaka H.,
RA    Tomasz A.;
RT    "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
RT    alanine amidase domain and an endo-beta-N-acetylglucosaminidase
RT    domain: cloning, sequence analysis, and characterization.";
RL    Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=NCTC 8325-4;
RA    Foster S.J.;
RL    Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC    -!- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOBIOSYL UNIT IN
CC        HIGH-MANNOSE GLYCOPEPTIDES AND GLYCOPROTEINS CONTAINING THE
CC        -[(MAN)5(GLCNAC)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
CC        RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
CC        OLIGOSACCHARIDE IS RELEASED INTACT.
CC    -!- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
CC        residues and L-amino acid residues in certain bacterial cell-wall
CC        glycopeptides.
CC    -!- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC        unit in high-mannose glycopeptides and glycoproteins containing
CC        the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC        residue remains attached to the protein; the rest of the
CC        oligosaccharide is released intact.
CC    -!- SUBCELLULAR LOCATION: Secreted.
CC    -!- PTM. UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
CC        EXTRACELLULAR LYTIC ENZYMES.
CC    -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
CC        ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
CC    -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
CC        GLYCOSYL HYDROLASES.
CC    ---------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute.  There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------------
DR    EMBL; D17366; BAA04185.1; -.
DR    EMBL; L41499; AAA99982.1; -.
DR    InterPro; IPR002502; Amidase_2.
DR    InterPro; IPR002901; Amidase_4.
DR    Pfam; PF01510; Amidase_2; 1.
DR    Pfam; PF01832; Amidase_2; 1.
DR    SMART; SM00644; Ami_2; 1.
DR    SMART; SM00047; LYZ2; 1.
KW    Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
FT    SIGNAL        1     29       POTENTIAL.
FT    CHAIN        30   1256       BIFUNCTIONAL AUTOLYSIN.
FT    DOMAIN      199    775       N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
FT    DOMAIN      776   1256       ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
FT    REPEAT      425    589       1.
FT    REPEAT      596    758       2.
FT    REPEAT      770    932       3.
SQ    SEQUENCE   1256 AA;  137384 MW;  2BB76CAA292FDD20 CRC64;
```

```
Qy    598  PTPKPSTPTTNNKLTVSSLNGV-----AQINAKNNGLFTTVYDKTGKPTKEVQKTFAVTKE 653
               |||.|||  |:||:   :|:      |:| |
Db     41  --GGTSF--AGHNSA----ICASSASYINFEGPQN-PDNYLVTPB---LSLPNG-- 82

Qy    654  ASLGGNKFYLVKDYNSPTLIGWVKQGDVIYNNAKSPVNVMQTYTVKPGTKLIYSVPWGTYK 713
               :||.       |                  ||| :  |||| ||:|    |||  ||:
Db     83  --------GTLTFWVCAQDANYASEHYAVIASSTGNDASNFAN-------ALLEVLT 125

Qy    714  QEAGAVSGTGNQTFKATKQQQIDKS----IYLFGTVNGKSGWVSKAYLAVPAAPKKAVAQ 769
               |  :   : |:  .:  :| |::       ||:|||||.|| | | |  |: |:
Db    126  AKT------VVTAPEAIR-------GTRVQGTWTQKTVQLPAGTKYVAFRHFGCT-- 167

Qy    770  PKTAVKAYTVTKPQTTQTVSKIAQVKPNNTGIRASVVEKTAK--NGAKY-ADRTFYVTKE 826
               |:| |:|| : | |   ||. : ||  | |:|.   |:|      |||   :|.|| |
Db    168  --------DFFWINLDDVEIKANGKRADFTE--TFESSTHGEAPAEWTTI 207

Qy    827  RAHGNEYVLLNNTSHNIPLGWFNVKDLNVQNLGKEVKTQKYTVNKSNNGLSMVPWGT- 885
               :| |||  |   || | || | |:  ||:|  |||||| ||| :| || |||
Db    208  DADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWN------GMALNPDNYLJSKDVTGAT 261

Qy    886  -----KNQVILTGNNI----AQGTFNATKQVSVGKDVYLYGTINNRTGWVNAKDLTAPT 935
                     |.:  |.:     | | |   :| |.| | |  || : ||:  | | |
Db    262  KVK--------YYYAVNDGFPGDHYAVMISKTGT------NAGDFTVVFEETPNG--- 302

Qy    936  AVKPTTSAAKDYNYTYVIKNG-NGYYYVTFNSDTAKYSLKAFNEQPFAVVKEQVINGQTW 994
               | |:|: | |:|   | :|| |  ||||  |||| || |  :|||  | :|| |  ||
Db    303  ---INKGGARFGLSTEADGAK----------PQSVWIERTVDLPAGTKY 338

Qy    995  YYGKLSNGKLAWIKST--DLAKELIKYNQTGMTLNQVAQIQAGLQY 1038
               |  :| || |.| |:    ||  |:  |::| :  | |:  | |
Db    995  YYGKLSNGKLAWIKST--DLAKELIKYNQTGMTLNQVAQIQAGLQY 1038
```

```
RESULT 10
PME_ASPAC
ID    PME_ASPAC      STANDARD;      PRT;    331 AA.
AC    Q12535;
DT    01-NOV-1997 (Rel. 35, Created)
DT    01-NOV-1997 (Rel. 35, Last sequence update)
DT    28-FEB-2003 (Rel. 41, Last annotation update)
DE    Pectinesterase precursor (EC 3.1.1.11) (Pectin methylesterase) (PE).
GN    PME1.
OS    Aspergillus aculeatus.
OC    Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC    Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX    NCBI_TaxID=5053;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=KSM 510;
RX    MEDLINE=97079238; PubMed=8920970;
RA    Christgau S., Kofod L.V., Halkier T., Andersen L.N., Hockauf M.,
RA    Dorreich K., Dalboege H., Kauppinen S.;
RT    "Pectin methyl esterase from Aspergillus aculeatus: expression
RT    cloning in yeast and characterization of the recombinant enzyme.";
RL    Biochem. J. 319:705-712(1996).
CC    -!- FUNCTION: Involved in maceration and soft-rotting of plant tissue.
CC    -!- CATALYTIC ACTIVITY: Pectin + N H(2)O = N methanol + pectate.
CC    -!- SIMILARITY: Belongs to the pectinesterase family.
CC    ---------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute.  There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------------
DR    EMBL; U49378; AAB42153.1; -.
DR    InterPro; IPR000070; Pectinesterase.
DR    Pfam; PF01095; Pectinesterase; 1
DR    PROSITE; PS00800; PECTINESTERASE_1; 1.
DR    PROSITE; PS05503; PECTINESTERASE_2; 1.
KW    Hydrolase; Aspartyl esterase; Cell wall; Signal.
FT    SIGNAL        1     17       BY SIMILARITY.
```

```
Query Match             4.8%;  Score 117.5; DB 1;  Length 1256;
Best Local Similarity  22.1%;  Pred. No. 1.8;
Matches 103;  Conservative 47; Mismatches 163;  Indels 153;  Gaps 26;

Qy      1  PNPNPGTTT----LSESFENGIPASWKTIDADGDG---NNWTTTPPP-----------  40
```

```
FT   CHAIN        18    331       PECTINESTERASE.
FT   ACT_SITE    162    162       BY SIMILARITY.
FT   ACT_SITE    183    183       BY SIMILARITY.
SQ   SEQUENCE    331 AA;  35681 MW;  1F1C81BF1E32174F CRC64;

Query Match             4.8%;  Score 116.5;  DB 1;  Length 331;
Best Local Similarity   22.9%;  Pred. No 0.36;
Matches   70;  Conservative  33;  Mismatches  98;  Indels 105;  Gaps  15;

Qy  102 AVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGT-KYVA 160
            :|:::| |:: | :| |: :|:    :|         |||||| :||
Db   27 AIVVAKSGDYTTIGDAI--DALSTSTTDTQTIFIE---EGT-YDEQVYLPAMTGKVII    79

Qy  161 FRHFGCTDFFWINL----DDVEIKANGKRADFTETFE---------SSTHGEAPAE   203
            :| : |  |   :|:|:|    |::||
Db   80 YGQTENTDSYADNLVTITHAISYEDAGESDDLTATFRNKAVGSQVYNLNIANTCGQACHQ 139

Qy  204 WTTIDADGDGQGWL-CLSSGQLDWLTAHGGTNVV-----------------       236
            : | |   |  |: |
Db  140 ALALSAWADQQGYYGCNFTGYQDTLLAQTGNQLYINSYIEGAVDFIFGQHARAWFQNVDI 199

Qy  237 -----------ASFSWNGMALNPDN-YLISKDVTGA------TKVKYYYAVNDGFPGDHYAV 280
                  |:|::
Db  200 RVVEGPTSASITANGRSSETDTSYYVINKSTVAAKEGDDVAEGTYYL----GRPWSEYA-  254

Qy  281 MISKTGTNAGDFTVVFEET--PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKY 338
             |||:|
Db  255 ------------------------RVVFQQTSMTNVINSLG                    284

Qy  339 VAFRHY 344
         | | :|
Db  285 VTFGEY 290

RESULT 11
C5AA_BACUD
ID   C5AA_BACUD   STANDARD;   PRT;   1385 AA.
AC   Q45760;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Pesticidal crystal protein crySAa (Insecticidal delta-endotoxin
DE   CryVA(a)) (Crystaline entomocidal protoxin) (152 kDa crystal protein).
GN   CRYSAA OR CRYVA(A) OR CRYVA.
OS   Bacillus thuringiensis (subsp. darmstadiensis).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=132264;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL B-18243 / PS17;
RA   Sick A.J., Schwab G.E., Payne J.M.;
RT   "Genes encoding nematode-active toxins cloned from Bacillus
RT   thuringiensis isolate PS17.";
RL   Patent number US5281530, 25-JAN-1994.
-!-  FUNCTION: ENDOTOXIN WITH NEMATICIDAL ACTIVITY.
-!-  DEVELOPMENTAL STAGE: The crystal protein is produced during
CC   sporulation and is accumulated both as an inclusion and as part of
CC   the spore coat.
-!-  MISCELLANEOUS: Toxic segment of the protein is located in the N-
CC   terminus.
-!-  SIMILARITY: Belongs to the delta endotoxin family.
```

```
DR   EMBL; L07025; AAA67694.1; -.
DR   PIR; T18213; T18213.
DR   InterPro; IPR005638; endotoxin_C.
DR   InterPro; IPR005639; endotoxin_N.
DR   InterPro; IPR008979; Gal_bind_like.
DR   Pfam; PF03944; endotoxin_C; 1.
DR   Pfam; PF03945; endotoxin_N; 1.
KW   Toxin; Sporulation.
SQ   SEQUENCE    1385 AA;  152439 MW;  E29FF11FC799DB95 CRC64;

Query Match             4.8%;  Score 116.5;  DB 1;  Length 1385;
Best Local Similarity   20.0%;  Pred. No 2.4;
Matches  119;  Conservative  67;  Mismatches 189;  Indels 219;  Gaps  30;

Qy    2 NPNPGTTLSESFENGI--------PASWKTIDADGDNNWTTTPPPG---GTSF-       45
         :|               |:  :||  :  | | :|  : :|||
Db  416 DPQAGPNVVSIDSSNPIIQINMDTWKTPPQGASG--WNTNLMRGSVSGLSFLQRDGTRLS 473

Qy   46 ---AGHNSAICASSA---SYINFEGPQNPDNY--------------LVTP-EISLPN--   81
            ||   :   ::|     | : |      |:             ||
Db  474 AGMGGGFADTIYSLPATHYLSYLLYGTPYQTSDNYSGHVGALVGVSTPQEATLPNIIGQPD 533

Qy   82 ------------GTLT-FWVCAQDA---------------NYASEHYAVYASSTG      109
                        | |:  | :|:|                 ::|
Db  534 EQGNVSTMGFPFEKASYGGTVVKEWLNGANAMKLSPGQSIGIPITNVTSGEYQIRCRYAS  593

Qy  110 NDASNF-----------ANALLEVLTAKTVVTAPEAIRGTRVQG------TWYQKTV    150
         || :|            ||   :                  | |||         |
Db  594 NDNTNVFFNVDTGGANPIFQQINFASTV-----DNNTGVQGANGVYVVKSIATTDNSFT  647

Qy  151 QLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDAD 210
         :|| ||     | :|  ::|::|:|::|       : ::|:  |||: :|  : ||:|
Db  648 EIPAKTINVHLTNQGSSDVF---LDRIEIPFSLPLIYHGSYNTSS-GADDVLW------  697

Qy  211 GDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISK------DVTGATKV  263
                                  :|  :: |:||: :|||::|||:|:
Db  698 ------SSSNMNYY-----DIIVNGQANSSSIASSMHLLNKGKVIKTIDIPGHSET    742

Qy  264 KY-YYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEADGAKPQ 322
         ::  |::   ||  :|::     |:  ||||:|     ||||||| ||| |:||| |
Db  743 FFATFPVPEGF--NEVRILAGLPEVSGNITVQSNNPPQPSNNGGGDGGNGGGDGGQ--  797

Qy  323 SWIERTVDLPAGTKYVAFRHYNCSDLNVILLDDIQFTMGGSPTPT------------  368
                                    :|::                | :
Db  798 ------------------------YN--------FSLSGSDHTTIYHGKLETGIHVQG 823

Qy  369 DYYYT-------VYRDGTKIKEGLTETTF------EBD-------GVATGNHEYCV    404
         :|::|        ::|:                  | :
Db  824 NYTYTGTPVLILNAYRNNTVVSSIPVYSPFDITIQTEADSLELELQPRYGFATVNGTATV 883

Qy  405 E---VKYTAGVS-PKECVNVT--VDPVQFNPVQNLTGSAVG---QKVTLKWDA      448
         :     ::|||   |::   :  :||| ::| || :  ::    | |:|:|||
Db  884 KSPNVNYDRSFKLPIDLQNITTQVNALFASGTQNMLAHNVSDHDIEEVVLKVDA      937

RESULT 12
YK28_ARCFU
ID   YK28_ARCFU   STANDARD;   PRT;   607 AA.
AC   O28251;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AF2028.
GN   AF2028.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
```

```
RA    Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA    Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA    Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA    Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA    Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA    Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA    Venter J.C.;
RT    "The complete genome sequence of the hyperthermophilic, sulphate-
RT    reducing archaeon Archaeoglobus fulgidus.";
RL    Nature 390:364-370(1997).
CC    -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC    -!- SIMILARITY: TO M.JANNASCHII MJ1393 AND MJ1394.
CC    --------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------
DR    EMBL; AE000963; AAB89227.1; -.
DR    PIR; C69503; C69503.
DR    TIGR; AF2028; -.
KW    Hypothetical protein; Transmembrane; Complete proteome.
FT    TRANSMEM      16     36       POTENTIAL.
FT    TRANSMEM     583    603       POTENTIAL.
SQ    SEQUENCE   607 AA;  66215 MW;  F46F16BF2AAFBE1 CRC64;

Query Match             4.7%;
Best Local Similarity  18.4%;   Score 114.5;  DB 1;  Length 607;
Matches   90;   Conservative   67;   Mismatches  187;  Indels  145;  Gaps   24;

Qy      3 PNPGTTTLSESFENGIPASWKTID-ADG-------------DGNNWTTTPPPGGTSFAGHN  49
            :  :|||||  ||| |::||  :|||             ||    |||:::
Db    201 PSSVTLITVNKNLNTSYDTNWKTLEFADGSSNPKQIISGTLDNNKWLTTTFKVKGNYSTDE 260

Qy     50 --SAICASSASYINFEGPQNPDNYLVTPELSLPNGTLTFWVCAQDANYASEHYAVYAS 106
            :    |||||  |||  || :|||||||||||||||:||:|  |||||||||||||
Db    261 AFVSRLVGFGEAYFQF-GPINGN-----------ISGTKIIDV-----FAIGNASIGVN 302

Qy    107 STGNDASN-FANALLEEVLTAKTVVTAPEAIRGT-----RVQGTWYQKTVQLPAGTKYV 159
            |||:|||   || ||||||| :|:|||| |||     ::|||||||||||||||||
Db    303 KSGPDNSNQWTGNVTIKNTATGLTYIVKSVKVWATDRNYNEINGARYENTTV------- 354

Qy    160 AFRHFGCTDFFWINLDDVEIKANGKRADFTETFESSTHGEAPAEWTTID---ADGDGQGW 216
                    :|| |  |||  | ||| |:|||||||||||||||||     ||||||||
Db    355 ---------------NVQI--GRDESFTSKDLSFQYDKVPIIWGNVTFRLVEDANYGW 395

Qy    217 LCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGD 276
            |||||||||||||||||  |:   :|| |  |||||:| |||| ||:|  ||| | |
Db    396 ---GVGQ-DKITDGGNTYIIERI-------YVIGSYLVKVTK-HVESAGNDIY----- 436

Qy    277 HYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLST-EADGAKPQ--SVWIERTVDL 332
            ||| |||:||| ||| ||  |||||||  || ||| |  ||||: |   :|| ||||||
Db    437 ----------NFSLTNGDNDWKDPQDRDGDWVNKSSML 489

Qy    333 PAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTFDYTYTV-----------YRDGTK 380
            ||||||||||:  ||  :| ||||||:|||||||||||||            |:| ||
Db    490 ---------------NIQL----SGYDIGYWWRIRPINASADGDGAYDDYTE 530

Qy    381 IKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDP------VQFNPV 430
            |::| |:  |  |  | :|  ||:  |:   ||: ||:|||         |:| |
Db    531 IENNQVVIFQ---IQGSEDYKLLDAFIVGIDPILSMNEQTSPKITLVSGAKATSYESV 587

Qy    431 QNLTGSAVG 439
            ||| |:  |
Db    588 MALATALVG 596
```

RESULT 13
YD94_META
ID   YD94_METUA   STANDARD;   PRT;   987 AA.

```
AC    Q58789;
DT    15-JUL-1998 (Rel. 36, Created)
DT    15-JUL-1998 (Rel. 36, Last sequence update)
DT    28-FEB-2003 (Rel. 41, Last annotation update)
DE    Hypothetical protein MJ1394.
GN    MJ1394.
OS    Methanococcus jannaschii.
OC    Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC    Methanocaldococcaceae; Methanocaldococcus.
OX    NCBI_TaxID=2190;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX    MEDLINE=96337999; PubMed=8688087;
RA    Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA    Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA    Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA    Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA    Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA    Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA    Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA    Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT    "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT    jannaschii.";
RL    Science 273:1058-1073(1996).
CC    -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC    -!- SIMILARITY: TO M.JANNASCHII MJ1393 AND A.FULGIDUS AF2028.
CC    --------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------
DR    EMBL; U67579; AAB99404.1; -.
DR    PIR; A64474; A64474.
DR    TIGR; MJ1394; -.
KW    Hypothetical protein; Transmembrane; Complete proteome.
FT    TRANSMEM      12     32       POTENTIAL.
FT    TRANSMEM     958    978       POTENTIAL.
SQ    SEQUENCE   987 AA;  112360 MW;  D1E628FFB28CA86D CRC64;

Query Match             4.6%;
Best Local Similarity  19.5%;   Score 112.5;  DB 1;  Length 987;
Matches   78;   Conservative   69;   Mismatches  143;  Indels  111;  Gaps   20;

Qy      7 TTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGG---------TSFAGHNSAICAS  55
           :|:|:                                        :::
Db    425 STSLNNDPNGGYKLLPKEIGRDVVISGWVYRPSNWGGGPIDRIGLEDENFDGYSFEVNHY 484

Qy     56 SASYINFEGPQNPDNYLVTPELSLPNGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 115
           :  :                                     :::
Db    485 S-NYISIDRRTNGNPTEISPEV---------YWNPPEDEWYFE-LKIYSNGTITFSTYY 533

Qy    116 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 175
               :                                       :::   ||
Db    534 QNGSL----AATVST----------IDNT-YTKFDRVVIHGGYVYY------------VD 566

Qy    176 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 235
                                                      :::
Db    567 DLEV--NSKNPDF------------YGDKNWKYLEITSANSSEGTAVLFDG--DYFKKDYNTSN 614

Qy    236 VASFSWNGMALNPDNVLISKDVTGATKV-----KYYAVND-----GFPGDHYAVMISK 284
            :::                                :::
Db    615 LNAINWTNITILNWSN------------DSATIVFNVLGNIYSYSERDNILAKYGFA--- 663

Qy    285 TGTNAGDFTVVFEETPNGINKGGAARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRH 343
           |||||                                           :::|||
Db    664 NGTNT-------NTSIKGVYASGS-YSISTDHGTTGEINIWIEN--------VTFKND 705

Qy    344 -----YNCSDLNYILLDDIQFTMGGSPTPDYTYTVYRDGT 379
```

Best Local Similarity 19.0%; Pred. No. 4.1;
Matches 94; Conservative 77; Mismatches 193; Indels 131; Gaps 26;

```
Qy   36  TTPPGGTSFAGHNSAICASSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDAN   95
             :: :::  :::     :  :  :  :   :  :  : :  :  :  :  :  :::
Db  437  TAEAPGNTKTVGYPYGILLSA---ISFDATR-----------------NGLALAAPALGQDVG  479

Qy   96  YASEHYA-----------------VYASST---GNDASNFAN---ALLEBVLTAKTV  129
             :  :::                  : : :   :  :  :  :::      : :
Db  480  Y---HFVPRLAVGGVSSPRGANGNIFLGSAITWGTNGGNFLDTKWHSPAVIEDAPTTFIT  536

Qy  130  VTAPEAIRGTRVQGTWYQKTVQLP-AGTKYVAFRHFGCTDFFWINLDDVEIK---ANGK  184
             : :: :  ::  ::  : : :   : : :  : ::    ::  ::  :  : :
Db  537  VNSSGVLQNS---GSQQSTSTPMPNSNGNESIPYRWTNSYDYNSVRFAALISKPAGGNTK  593

Qy  185  RAD--FTETFESSTHGEAPAEWTTID------ADGDGQGWICLSSGQLDWLTAHGGTNVV  236
             ::   : :  :  :  :::  :::         : : :  : :  : :::  ::::  :::
Db  594  QVESLFTTALKLDTLNSLPNKFTQENNIFFSYAMLDGRQWSLGTRKDSTWLT---TNTI  649

Qy  237  ASFSWN---------------GMALNPDNYLIS----------------KDV---TGATKVKYYAV  269
             ::::                :  :  :  :                :                  : :::: ::
Db  650  NNFTYNTQQOLASTAAGENANPRNILNALITTAKGFDRRDIGNVDILYSNNTNKFTYYQV  709

Qy  270  ND----------GFPGDHYAVMISKTGTNAGDFYVFEETPNGINK-GGARFGLSTEADG  318
               :            :::  :  :  :  :  :  :  :::  :::  :::  ::::::
Db  710  GGAITTWPEVQVNYKTSANITYYNLTRIDFGSTTPATQDANTVSSKLNGAYLSSTGDQQG  769

Qy  319  AKPQSVWIERTVDLPAGTKYV--APRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYR  376
             : :: :    :  :  : ::     ::  :::  :::  ::  :::  :  :  :  :::
Db  770  WYNGSIYVKKASFTPSSQGYTWQDFKGLTTASNAVISN---WTKAG------YSIRP  818

Qy  377  DGTKIKEGLTETFEEDGVATGN---HEYCVE------VKYTAGVSPKBECVNTVDPVQ  426
             ::  :::   :  :  :::  :::           : :::  :::  : : :  : :::
Db  819  DDDTV-FSVSKIPFEKEITAAVNVRSLDSYYVQLNGETSVNTVARVSFDSSA-LTLNPKR  876

Qy  427  F-NPVQNLTGSAVGQ  440
             :  ::::::  :::
Db  877  ITNPLMN-RDNVIGQ  890
```

```
Db  706  AKSYSFNLTNLNIWAVNKSAYELYWNP----FNKSIWIDGS  742
             :  :::      : :  :  ::   :::        ::: ::
```

RESULT 14
ADP1_MYCGA
ID   ADP1_MYCGA   STANDARD;   PRT;   1122 AA.
AC   Q49379; Q49437; Q53351;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Adhesin P1 precursor (Cytadhesin P1) (Attachment protein).
GN   GAPA OR MGC1.
OS   Mycoplasma gallisepticum.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2096;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S6;
RX   MEDLINE=96201559; PubMed=8613358;
RA   Keeler C.L. Jr., Hnatow L.L., Whetzel P.L., Dohms J.E.;
RT   "Cloning and characterization of a putative cytadhesin gene (mgc1)
RT   from Mycoplasma gallisepticum.";
RL   Infect. Immun. 64:1541-1547(1996).
RN   [2]
RP   SEQUENCE OF 1-12 FROM N.A.
RC   STRAIN=S6;
RA   Hnatow L.L., Keeler C.L. Jr., Tessmer L., Dohms J.E.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 378-570 FROM N.A.
RC   STRAIN=S6;
RX   MEDLINE=93371270; PubMed=8363503;
RA   Dohms J.E., Hnatow L.L., Whetzel P., Morgan R., Keeler C.L. Jr.;
RT   "Identification of the putative cytadhesin gene of Mycoplasma
RT   gallisepticum and its use as a DNA probe.";
RL   Avian Dis. 37:380-388(1993).
RN   [4]
RP   SEQUENCE OF 159-1122 FROM N.A.
RC   STRAIN=S6;
RA   Goh M.S., Geary S.J.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Could be involved in cytadherence.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the adhesin P1 family.
CC   -!- CAUTION: Ref 4 sequence differs from that shown due to a
CC   frameshift in position 159 to 213.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U34842; AAB02987.1; -.
DR   EMBL; U44804; AAC83385.1; ALT_FRAME.
DR   PIR; T18346; T18346.
KW   Cytadherence; Signal; Transmembrane.
FT   SIGNAL   1   30   POTENTIAL.
FT   CHAIN   31   1122   ADHESIN P1.
FT   TRANSMEM   997   1021   POTENTIAL.
FT   DOMAIN   1002   1009   POLY-ILE.
FT   CONFLICT   313   314   DM -> IW (IN REF. 4).
FT   CONFLICT   382   382   Y -> F (IN REF. 3).
FT   CONFLICT   501   501   I -> T (IN REF. 3).
FT   CONFLICT   568   568   R -> G (IN REF. 3).
FT   CONFLICT   570   570   T -> A (IN REF. 3).
FT   CONFLICT   693   695   DIL -> VYT (IN REF. 4).
SQ   SEQUENCE 1122 AA; 121305 MW; 155C34DA2D6C3C65 CRC64;

Query Match   4.6%; Score 111.5; DB 1; Length 1122;

RESULT 15
BGAL_BACME
ID   BGAL_BACME   STANDARD;   PRT;   1034 AA.
AC   O52847;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Beta-galactosidase (EC 3.2.1.23) (Lactase).
GN   BGAM.
OS   Bacillus megaterium.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1404;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 319;
RA   Strey J.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC   galactose residues in beta-D-galactosides.
CC   -!- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AJ000733; CAA04267.1; -.
DR   PIR; T30574; T30574.
DR   HSSP; P00722; 1BGL.
DR   InterPro; IPR008979; Gal_bind_like.

```
DR   InterPro; IPR006101; Glyco_hydro_2.
DR   InterPro; IPR006102; Glyco_hydro_2Ig.
DR   InterPro; IPR006104; Glyco_hydro_2SB.
DR   InterPro; IPR006103; Glyco_hydro_2TIM.
DR   InterPro; IPR004200; Glyco_hydro_42C.
DR   InterPro; IPR004199; Glyco_hydro_42N.
DR   Pfam; PF02930; Bgal_small_C; 1.
DR   Pfam; PF02929; Bgal_small_N; 1.
DR   Pfam; PF00703; Glyco_hydro_2; 1.
DR   Pfam; PF02836; Glyco_hydro_2_C; 1.
DR   Pfam; PF02837; Glyco_hydro_2_N; 1.
DR   PRINTS; PR00132; GLHYDRLASE2.
DR   PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR   PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KW   Hydrolase; Glycosidase.
FT   ACT_SITE       481     481      PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE       547     547      NUCLEOPHILE (BY SIMILARITY).
SQ   SEQUENCE   1034 AA;  118673 MW;  3B644C9A649415E9 CRC64;
```

```
Query Match             4.5%;  Score 110.5;  DB 1;  Length 1034;
Best Local Similarity  19.9%;  Pred. No. 4.3;
Matches  114;  Conservative   70;  Mismatches  169;  Indels  221;  Gaps   35;

Qy      6 GT-TTLSESFENGIPAS----WK--TIDA-----DGDGNN----WTTTPPGGTSFAGHN 49
            ||  |     |||        : :  ||       :  |  :       |   ||    ||
Db    432 GTWTIIQEGEQKAVPGSKPEWKENVLDRCRSMYERDKNHPSIIIWSL----GNESFGGEN 487

Qy     50 SAIC-----ASSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAV 103
            :  |        ||:| |     :  :: :      |  |   |::|   || :
Db    488 FQHMYTFFKEKDSTRLVHYEGLFHHRDYDASDIES------TMVYKPADV---ERYAL 536

Qy    104 --------YASSTGNDASNFAN--ALLEEVLTAKTVVTAPEAIRGTRVQG----TWY 146
                      :           |  |  |     |      ||
Db    537 MNPKKPYILCEYSHAMGNSCGNLYKYWELFDQY------------PIIQGGFIWDWK 581

Qy    147 QKTVQLPA--GTKYVAF-RHFGCT----DFFWINLDDVEIKANGKRADFTETFESSTHGE 199
            :  :|:       :|:  |  ::      | :    :  ||   |  |  |:  :  :
Db    582 DQALQATAEDGTSYLAYGGDFGDTPNDGNFCGNGLIFADGTASPKIAEVKKCYQ------ 635

Qy    200 APAEWTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVA-SFSW------NG------MA 245
            : |:| |     || ||  :     :| |  | : |   ||        :
Db    636 -PVKWTAVDP-------------AKGKFAVQNKHLFTNLNAYDFVWTVEKNGELVEKHASLLN 684

Qy    246 LNPD-----------------------NYLISKDVTGAT---KVKY-- 265
            :|                          |||  |  :      ||||
Db    685 VAPDGTDELTLSYPLYEQENETDEFVLTLSLRLSKDTAWASAGYEVAYEQFVLPAKAAMP 744

Qy    266 -YYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTE-----ADGA 319
             :  :|       :: ||||        : |:::     :|        |        :
Db    745 SVKAAHPALTVDQNEQTLTVTGTN---FTAIFDK------RKGQFISYNYERTELLASGF 795

Qy    320 KPQSVWIERTVDLPAGTKYVAFRHYNCS---------------DLNYILLDDIQFTM 361
            :  :                 :         ::::: :: :|
Db    796 RP-NFW-RAVTDNDLGNKL---HERCQTWRQASLEQHVKKVTVQPQVDFVII-SVELAL 848

Qy    362 GGSPTPTDYYTVYRDG-TKIKEGL--TETTFE---------------- 391
            || :| |: || |||   | ||::   :|:|:|
Db    849 DNSLASCYVTYTLYNDGEMKIEQSLAPSETMPEIPEIGMLFTMNAAFDSLTWYGRGPHEN 908

Qy    392 ----EDGVATGNHEYCVEVKYTAGVSPKECVNVT 421
             :        :  :  ::: | || |  :  |||
Db    909 YWDRKTGAKLALHKGSVKEQVTPYLRPQECGNKT 942
```

Search completed: May 18, 2004, 11:43:32
Job time : 10.4245 secs

```
                GenCore version 5.1.6
           Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        May 18, 2004, 11:36:20 ; Search time 32.8982 Seconds
                                         (without alignments)
                                         4315.838 Million cell updates/sec

Title:         US-08-570-311-16
Perfect score: 2443
Sequence:      1 PNPNPGTTLSESFENGIPA.........QNLTGSAVGQKVTLKWDAPN 450

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:     1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 listing first 45 summaries

Database :     SPTREMBL_25:*
          1: sp_archea:*
          2: sp_bacteria:*
          3: sp_fungi:*
          4: sp_human:*
          5: sp_invertebrate:*
          6: sp_mammal:*
          7: sp_mhc:*
          8: sp_organelle:*
          9: sp_phage:*
         10: sp_plant:*
         11: sp_rodent:*
         12: sp_virus:*
         13: sp_vertebrate:*
         14: sp_unclassified:*
         15: sp_rvirus:*
         16: sp_bacteriap:*
         17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2064	84.5	1723	2	P72197	P72197 porphyromon
2	2063	84.4	1687	2	Q9R9B7	Q9R9B7 porphyromon
3	2063	84.4	1704	2	Q51816	Q51816 porphyromon
4	2051	84.0	1706	2	Q51839	Q51839 porphyromon
5	2049	83.9	1723	2	P72194	P72194 porphyromon
6	2041	83.5	1358	2	P96967	P96967 porphyromon
7	2040	83.5	1706	2	Q51838	Q51838 porphyromon
8	2031.5	83.2	1223	2	Q92ZN5	Q92ZN5 porphyromon
9	2016.5	82.5	1223	2	P72196	P72196 porphyromon
10	2016.5	82.5	1097	2	O07442	O07442 porphyromon
11	2011.5	82.3	1732	2	Q51817	Q51817 porphyromon
12	1997.5	81.8	1732	2	O52050	O52050 porphyromon
13	683	28.0	925	2	Q9F4J0	Q9F4J0 porphyromon
14	561.5	23.0	312	2	Q9KIB3	Q9KIB3 porphyromon
15	233	9.5	293	2	Q9XBU9	Q9XBU9 porphyromon
16	152.5	6.2	1742	16	Q8P377	Q8P377 xanthomonas

17	140	5.7	3346	16	Q7WN54	Q7WN54 bordetella
18	135	5.5	2215	16	Q7WBN0	Q7WBN0 bordetella
19	135	5.5	3988	17	Q8TPZ1	Q8TPZ1 methanosarc
20	134	5.5	1541	16	Q7UQJ9	Q7UQJ9 rhodopirell
21	133	5.4	2219	16	Q88WI9	Q88WI9 lactobacill
22	132.5	5.4	1115	4	Q9HD43	Q9HD43 homo sapien
23	132.5	5.3	1341	16	Q8UAU1	Q8UAU1 agrobacteri
24	130.5	5.3	2468	16	Q912M3	Q912M3 pseudomonas
25	129.5	5.3	756	9	Q858B6	Q858B6 enterobacte
26	129.5	5.3	1386	17	Q8TI72	Q8TI72 methanosarc
27	129.5	5.3	2656	5	Q9GNU3	Q9GNU3 paracentrot
28	129	5.3	7716	16	Q7UWZ8	Q7UWZ8 rhodopirell
29	128.5	5.3	680	2	Q52644	Q52644 ruminococcu
30	128.5	5.3	2230	16	Q7U7J7	Q7U7J7 synechococc
31	128.5	5.3	2768	16	Q8E9G6	Q8E9G6 shewanella
32	128	5.2	1672	16	Q8Y366	Q8Y366 ralstonia s
33	127.5	5.2	1684	2	Q03658	Q03658 unidentifie
34	127	5.2	1800	2	Q91948	Q91948 pseudomonas
35	127	5.2	8682	16	Q8RG2	Q8RG2 pseudomonas
36	126.5	5.2	1095	2	O85151	O85151 photorhabdu
37	126.5	5.2	3064	16	Q82XT8	Q82XT8 nitrosomona
38	126	5.2	1608	17	Q8PVI0	Q8PVI0 methanosarc
39	126	5.2	1649	16	Q9CFA2	Q9CFA2 lactococcus
40	126	5.2	2522	16	Q8EKA6	Q8EKA6 shewanella
41	125.5	5.1	555	16	Q826W1	Q826W1 streptomyce
42	125.5	5.1	955	17	Q8TQ91	Q8TQ91 methanosarc
43	125.5	5.1	5291	16	Q8X2T1	Q8X2T1 escherichia
44	124.5	5.1	691	16	Q9RZS7	Q9RZS7 deinococcus
45	124.5	5.1	1357	17	Q8THC8	Q8THC8 methanosarc

ALIGNMENTS

RESULT 1
P72197
```
ID   P72197        PRELIMINARY;       PRT;   1723 AA.
AC   P72197;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA   Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT   "Molecular cloning and characterization of Porphyromonas gingivalis
RT   Lys-gingipain.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U54691; AAA99910.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1723 AA;  186831 MW;  4508A7E50197CBBD CRC64;
```

Query Match   84.5%;   Score 2064; DB 2; Length 1723;
Best Local Similarity   86.3%;   Pred. No. 3.1e-127;

Query Match   84.4%;   Score 2063; DB 2; Length 1687;
Best Local Similarity   86.3%;   Pred. No. 3.5e-127;
Matches 389; Conservative 17; Mismatches 39; Indels   6; Gaps   4;

```
Qy      1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-   59
          ||||||||||||||||||||||||||||||||||||||:|||:||   ||||::|:
Db    939 PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG  995

Qy     60 INFEGPQNPDNLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
          :|||||| |||||||||||:|:||||||||||||||||||||||||||||||||:||||
Db    996 LGGIGVLTPDNVLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1055

Qy    120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVARFHGCTDFFWINLDDVEI  179
          ||||:|||  |:|||||||:||||||||||||||||||||||||:||||||||||||||
Db   1056 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLLDEVEI 1113

Qy    180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF  239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1114 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1173

Qy    240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1174 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1233

Qy    300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1234 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1293

Qy    360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1294 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1353

Qy    420 VTVDPVQFNPVQNLTGSAVGQXVTLKWDAPN  450
          ||::|  |||||||  ||||| |:||||||:
Db   1354 VTINPTQFNPVXNLKAQPDGDGDVVLKWEAPS 1384
```

RESULT 3
```
Q51816
ID   Q51816           PRELIMINARY;        PRT;   1704 AA.
AC   Q51816;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arg-gingipain-1 proteinase.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95138080; PubMed=7836351;
RA   Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
RA   Travis J., Barr P.J.;
RT   "Molecular cloning and structural characterization of the Arg-
RT   gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a
RT   proteinase-adhesin polyprotein.";
RL   J. Biol. Chem. 270:1007-1010(1995).
DR   EMBL; U15282; AAA69539.1; -.
DR   PIR; A55426; A55426.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
```

Matches 389; Conservative 17; Mismatches 39; Indels   6; Gaps   4;

```
Qy      1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-   59
          ||||||||||||||||||||||||||||||||||||||:|||:||   ||||::|:
Db    975 PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1031

Qy     60 INFEGPQNPDNLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
          :|||||| |||||||||||:|:||||||||||||||||||||||||||||||||:||||
Db   1032 LGGIGVLTPDNVLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1091

Qy    120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVARFHGCTDFFWINLDDVEI  179
          ||||:|||  |:|||||||:||||||||||||||||||||||:||||||||||||||||
Db   1092 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1149

Qy    180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF  239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1150 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1209

Qy    240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1210 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1269

Qy    300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1270 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1329

Qy    360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1330 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVN 1389

Qy    420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  450
          ||::|  |||||||   ||||:||||||:
Db   1390 VTINPTQFNPVXNLKAQPDGGDVVLKWEAPS 1420
```

RESULT 2
```
Q9R9B7
ID   Q9R9B7           PRELIMINARY;        PRT;   1687 AA.
AC   Q9R9B7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin/protease.
GN   HAGE.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Dong H., Progulske-Fox A.;
RT   "Cloning and characterization of hagE from P. gingivalis 381.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF026946; AAD01810.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE   1687 AA;  183702 MW;  D085B516A399FE70 CRC64;
```

```
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   CHAIN        228    719       MATURE 50-KDA CYSTEINE PROTEINASE
FT                                 GINGIPAIN.
SQ   SEQUENCE   1704 AA;  185436 MW;  6A34B40131C22A676 CRC64;

Query Match             84.4%;  Score 2063;  DB 2;  Length 1704;
Best Local Similarity   86.3%;  Pred. No. 3.5e-127;
Matches  389;  Conservative  17;  Mismatches  39;  Indels  6;  Gaps  4;

Qy      1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASY- 59
          ||||||||||||||||||||||||||||||||||       |:|| |||||||
Db    956 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1012

Qy     60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          : |||  ||||||::| :|||:|| ||||||||||||||||||||||||||||||||
Db   1013 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1072

Qy    120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||: ||  : :|||:||| | :|||||||||||||||||||||| |::|| ||||||
Db   1073 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1130

Qy    180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1131 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1190

Qy    240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1191 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1250

Qy    300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYHYNCSDLNYILLDDIQF 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1251 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYHYNCSDLNYILLDDIQF 1310

Qy    360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1311 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1370

Qy    420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          |||:|| :|::||:|||||
Db   1371 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401

RESULT 4
Q51839
ID   Q51839      PRELIMINARY;     PRT;  1706 AA.
AC   Q51839; Q51840;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arginine-specific thiol protease precursor.
GN   PRTR.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96160709; PubMed=7857299;
RA   Kirszbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N.,
RA   Reynolds E.C.;
RT   "Complete nucleotide sequence of a gene prtR of Porphyromonas
RT   gingivalis W50 encoding a 132 kDa protein that contains an arginine-
RT   specific thiol endopeptidase domain and a haemagglutinin domain.";
RL   Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96311339; PubMed=8713096;
RA   Slakeski N., Cleal S.M., Reynolds E.C.;
```

```
RT   "Characterization of a Porphyromonas gingivalis gene prtR that encodes
RT   an arginine-specific thiol proteinase and multiple adhesins.";
RL   Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Reynolds E.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Slakeski N.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L26341; AAC18876.1; -.
DR   HSSP; P95493; 1CVR.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        228   1706       ARGININE-SPECIFIC THIOL PROTEASE.
SQ   SEQUENCE   1706 AA;  185626 MW;  E8BDF07C9813B844 CRC64;

Query Match             84.0%;  Score 2051;  DB 2;  Length 1706;
Best Local Similarity   85.8%;  Pred. No. 2.2e-126;
Matches  387;  Conservative  18;  Mismatches  40;  Indels  6;  Gaps  4;

Qy      1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASY- 59
          ||||||||||||||||||||||||||||||||||       :|:|| |||||||
Db    958 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1014

Qy     60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          : |||  ||||||::| :|||:|| ||||||||||||||||||||||||||||||||
Db   1015 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1074

Qy    120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||: ||  : :|||:||| | :|||||||||||||||||||||| |::|| ||||||
Db   1075 LEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1132

Qy    180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSF 1192

Qy    240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1252

Qy    300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYHYNCSDLNYILLDDIQF 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1253 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYHYNCSDLNYILLDDIQF 1312

Qy    360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1313 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVN 1372

Qy    420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          |||:|| :|::||:|||||
Db   1373 VTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 5
P72194
```

ID   P72194,   PRELIMINARY;   PRT;   1723 AA.
AC   P72194;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97044756; PubMed=8889827;
RA   Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
RT   "Cloning and sequencing of the gene encoding a novel lysine-specific
RT   cysteine proteinase (Lys-gingipain) in Porphyromonas gingivalis:
RT   structural relationship with the arginine-specific cysteine proteinase
RT   (Arg-gingipain).";
RL   J. Biochem. 120:398-406(1996).
DR   EMBL; D83258; BAA11870.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0005508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; FN_III.
DR   InterPro; IPR005536; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1723 AA;   187261 MW;   562896 3D251493EB CRC64;

Query Match   83.9%; Score 2049; DB 2; Length 1723;
Best Local Similarity   85.8%; Pred. No. 3e-126;
Matches 387; Conservative 17; Mismatches 41; Indels 6; Gaps 4;

```
Qy     1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-   59
Db   975 PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1031

Qy    60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
Db  1032 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1091

Qy   120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
Db  1092 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1149

Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASF   239
Db  1150 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGTNVVASF 1209

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  299
Db  1210 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1269

Qy   300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
Db  1270 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1329

Qy   360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
Db  1330 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVN 1389

Qy   420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  450
Db  1390 VTINFTQFNPVKNLKAQPDGGDVVLKWEAPS 1420
```

RESULT 6
P96967

ID   P96967,   PRELIMINARY;   PRT;   1358 AA.
AC   P96967;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin.
GN   HAGD.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Lepine G., Whitlock J., Wojciechowski L., Progulske-Fox A.;
RT   "Cloning, sequencing and characterization of hagD, a member of the
RT   HArep multigene family in Porphyromonas gingivalis.";
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U68468; AAB49691.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0005508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; FN_III.
DR   InterPro; IPR005536; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   SMART; SM00060; FN3; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1358 AA;   147102 MW;   47FCA0B25B06DED8 CRC64;

Query Match   83.5%; Score 2041; DB 2; Length 1358;
Best Local Similarity   85.6%; Pred. No. 7.4e-126;
Matches 386; Conservative 17; Mismatches 42; Indels 6; Gaps 4;

```
Qy     1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-   59
Db   610 PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG  666

Qy    60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
Db   667 LGGIGVLTPDNYLITPALDLLANGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL  726

Qy   120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
Db   727 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI  784

Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASF   239
Db   785 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGTNVVASF  844

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  299
Db   845 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  904

Qy   300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
Db   905 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  964

Qy   360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
Db   965 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVN 1024

Qy   420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  450
```

Db   1025 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1055

RESULT 7
Q51838
ID   Q51838   PRELIMINARY;   PRT;   1706 AA.
AC   Q51838;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Protease precursor.
GN   PRPR1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96071894; PubMed=7591131;
RA   Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.;
RT   "Characterization, genetic analysis, and expression of a protease
RT   antigen (PrpRI) of Porphyromonas gingivalis W50.";
RL   Infect. Immun. 63:4744-4754 (1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.;
RT   "The prpR1 and the prR2 arginine-specific protease genes of
RT   Porphyromonas gingivalis W50 produce five biochemically distinct
RT   enzymes.";
RL   Mol. Microbiol. 23:0-0(1997).
DR   EMBL; X82680; CAA57997.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Signal.
FT   SIGNAL   1   23   POTENTIAL.
FT   CHAIN   228   719   ALPHA-PROTEASE.
FT   CHAIN   720   1262   BETA-ADHESIN.
SQ   SEQUENCE   1706 AA; 185705 MW; 0E56DCD87FDA8CDD CRC64;

Query Match   83.5%; Score 2040; DB 2; Length 1706;
Best Local Similarity   85.6%; Pred. No. 1.2e-125;
Matches 386; Conservative 17; Mismatches 42; Indels 6; Gaps 4;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASY- 59
          ||||||||||||||||||||||||||||||||||||||||||||| : || |||:
Db   958 PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1014

Qy   60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          |||||||||||||| |||| |||||| |||||||||||||||||||||||||||||||||
Db   1015 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNAL 1074

Qy   120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||: ||||||| :|||||| : :||| |:|| ||||||||||||||||| |||:||||:|
Db   1075 LEETITAKG-VRSPEAIRG-RIQSTWRQKIVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1132

Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGDGQWLCLSSGQLDWLTAHGGTNVVASF 239

Db   1133 KANGKRADFTETFESSTHGEATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSF 1192

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1252

Qy   300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1253 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1312

Qy   360 TMGGSPTPTDYTYTYVRDGTKIKEGLTETTFBEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1313 TMGGSPTPTDYTYTYVRDGTKIKEGLTETTFBEDGVATGNHEYCVEVKYTAGVSPKKCVN 1372

Qy   420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          ||| ||||||||||:||||||||:||:||:|
Db   1373 VTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 8
Q9ZNB5
ID   Q9ZNB5   PRELIMINARY;   PRT;   1223 AA.
AC   Q9ZNB5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   130k-HMGD (Fragment).
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=99143166; PubMed=9988746;
RA   Shibata Y., Hayakawa M., Takiguchi H., Shiroza T., Abiko Y.;
RT   "Determination and characterization of the hemagglutinin-associated
RT   short motifs found in Porphyromonas gingivalis multiple gene
RT   products.";
RL   J. Biol. Chem. 274:5012-5020(1999).
DR   EMBL; AB019363; BAA34341.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   NON_TER   1   1
SQ   SEQUENCE   1223 AA; 131542 MW; 00225CD2BA9F91B3 CRC64;

Query Match   83.2%; Score 2031.5; DB 2; Length 1223;
Best Local Similarity   85.6%; Pred. No. 2.7e-125;
Matches 387; Conservative 15; Mismatches 41; Indels 9; Gaps 6;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASY- 59
          ||||||||||||||||||||||||||||||||||||||||||||| : || |||:
Db   468 PNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 523

Qy   60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          |||||||||||||| |||| |||||| |||||||||||||||||||||||||||||||||
Db   524 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNFTNAL 583

Qy   120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||: ||||||| :|||||| : :||| |:|| ||||||||||||||||| |||:||||:|
Db   584 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 641

```
Db   398 LGGIGVLITPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 457

Qy   120 LEEVLTAKTVVIAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
         ||.:|..:|:||||::|||:.|:.:||:|||:.:::.||||:||||.||:.||:.|||:|:
Db   458 LEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 515

Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
         |||||||||||||||||||||||||||||||||||||||||||||:||||||:|||||||
Db   516 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVVSSF 575

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   576 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 635

Qy   300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   636 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 695

Qy   360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   696 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVN 755

Qy   420 VTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 449
         ||:|.|||||||||||.||  |||||||||||
Db   756 VTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 787
```

RESULT 10
```
007442
ID   O07442                     PRELIMINARY;            PRT;    1732 AA.
AC   O07442;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine-specific cysteine proteinase.
GN   PRTK.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=99235907; PubMed=10219167;
RA   Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtK that encodes
RT   a lysine-specific cysteine proteinase and three sequence-related
RT   adhesins.";
RL   Oral Microbiol. Immunol. 14:92-97(1999).
DR   EMBL; U75366; AAB60809.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25_C; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1732 AA;  187914 MW;  45D5B9137739I703 CRC64;

Query Match             82.5%;  Score 2016.5;  DB 2;  Length 1732;
Best Local Similarity   85.0%;  Pred. No. 4.2e-124;
Matches  384;  Conservative  18;  Mismatches  41;  Indels   9;  Gaps     6;

Qy     1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY- 59
         ||||.|.:|||||||||||||||||||||||||||||||||||||||||||||||:|||
Db   977 PNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1032
```

```
Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
         |||||||||||||||||||||||||||||||||||||||||||||:||||||:|||||||
Db   642 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 701

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   702 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 761

Qy   300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   762 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 821

Qy   360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   822 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVN 881

Qy   420 VTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 449
         ||:||.||||||||||.||  |||||||||||
Db   882 VTINPTQFNPVQNLTAEQAPNSMDAILKWNAP 913
```

RESULT 9
```
P72196
ID   P72196                     PRELIMINARY;            PRT;    1097 AA.
AC   P72196;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   TonB-linked adhesin precursor.
GN   TLA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97386416; PubMed=9244265;
RA   Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA   Curtis M.A.;
RT   "The tla gene of Porphyromonas gingivalis W50: a homologue of the
RT   arginine-specific protease precursor (PrpRI) which shares sequence
RT   similarity to TonB-linked receptors.";
RL   J. Bacteriol. 179:4778-4788(1997).
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR   EMBL; Y07618; CAA68897.1; -.
DR   GO; GO:0019867; C:outer membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Membrane; Outer membrane; Receptor; Signal; TonB box.
FT   SIGNAL       1     53       POTENTIAL.
SQ   SEQUENCE   1097 AA;  118731 MW;  73BBA337B421F8B9 CRC64;

Query Match             82.5%;  Score 2016.5;  DB 2;  Length 1097;
Best Local Similarity   85.0%;  Pred. No. 2.3e-124;
Matches  384;  Conservative  18;  Mismatches  41;  Indels   9;  Gaps     6;

Qy     1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY- 59
         ||||.|.:|||||||||||||||||||||||||||||||||||||||||||||:|||
Db   342 PNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 397

Qy    60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
         :.|.||..||||.||||.|.|||||:|||||||||||||||||||||||||||||||||||
```

```
Qy    60  INFEGPQNPDNYLIVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
              :||| |.:|||:|| |||||| |||||||||||||||||||||||||||||||.|||
Db  1033  LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL  1092

Qy   120  LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
          |||.:|||| :|:||:| : ||||||||||| |||||||||||||||| ||.:::|:|||
Db  1093  LEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI  1150

Qy   180  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF  239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||.||.||
Db  1151  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSF  1210

Qy   240  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1211  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  1270

Qy   300  PNGINKGGARPGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
          |||||||||.||||||| |||||||||||||||||||||||||||||||||||||||||
Db  1271  PNGINKGGARPGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  1330

Qy   360  TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||.||
Db  1331  TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVN  1390

Qy   420  VTVDPVQFNPVQNLTGSAV-GQKVTLKWDAP  449
          |||: ||||||||||  ||  |||||||.||
Db  1391  VTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP  1422
```

RESULT 11
Q51817
```
ID   Q51817      PRELIMINARY;      PRT;   1732 AA.
AC   Q51817;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Porphypain.
GN   PRTP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W12;
RX   MEDLINE=96213011; PubMed=8631659;
RA   Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA   Progulske-Fox A., Lantz M.S.;
RT   "Analysis of the prtP gene encoding porphypain, a cysteine proteinase
RT   of Porphyromonas gingivalis.";
RL   J. Bacteriol. 178:2734-2741(1996).
DR   EMBL; U42210; AAB06565.1; -.
DR   PIR; T30836; T30836.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase C25 C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1732 AA;  187875 MW;  65427IDBEF7BCAE4 CRC64;
```

Query Match   82.3%; Score 2011.5; DB 2; Length 1732;
Best Local Similarity 84.7%; Pred. No. 9e-124;
Matches 383; Conservative 19; Mismatches 41; Indels 9; Gaps 6;

```
Qy     1  PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-   59
          ||||||.||||||||||||||||||||||||  |  | |||.||||||.  : :|:: :
Db   977  PNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG  1032

Qy    60  INFEGPQNPDNVLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
          : |  |.:|||:|| |||||| |||||||||||||||||||||||||||||||.|||
Db  1033  LGGIGVLTPDNVLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL  1092

Qy   120  LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
          |||.:|||| :|:||:| : ||||||||||| |||||||||||||||| ||.:::|:|||
Db  1093  LEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI  1150

Qy   180  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF  239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||.||.||
Db  1151  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSF  1210

Qy   240  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1211  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  1270

Qy   300  PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  1271  PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  1330

Qy   360  TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||.||:
Db  1331  TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVD  1390

Qy   420  VTVDPVQFNPVQNLTGSAV-GQKVTLKWDAP  449
          |||: ||||||||||  ||  |||||||.||
Db  1391  VTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP  1422
```

RESULT 12
O52050
```
ID   O52050      PRELIMINARY;      PRT;   1732 AA.
AC   O52050;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine specific cysteine protease.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=98298016; PubMed=9632563;
RA   Lewis J.P., Macrina F.L.;
RT   "IS195, an insertion sequence-like element associated with protease
RT   genes in Porphyromonas gingivalis.";
RL   Infect. Immun. 66:3035-3042(1998).
DR   EMBL; AF017059; AAC26523.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25 C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE   1732 AA;  187931 MW;  B2337463D5CB5EA5 CRC64;
```

Query Match   81.8%; Score 1997.5; DB 2; Length 1732;

```
Best Local Similarity   84.3%;  Pred. No. 7.5e-123;
Matches  381;  Conservative  19;  Mismatches  43;  Indels  9;  Gaps  6;

Qy      1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-
          |||||||| :||||||||||||||||||||||||||||||| |:|||||||||||||||:
Db    977 PNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1032

Qy     60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          :  :|||||||||| :||||| ||| | ||||||||:||||||||||||||||||||||
Db   1033 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1092

Qy    120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||| :|||| :|||||||||||:|:||| ||||||||||||||||||||||||||||||
Db   1093 LEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1150

Qy    180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db   1151 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSF 1210

Qy    240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1270

Qy    300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1330

Qy    360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          |||||||||||||:||||||||||||:||:||||:||||||||||||||||||||:|||
Db   1331 TMGGSPTPTDYTYTVYRDGTNIKEGLTTTFEYGVATGNHEYCVEVKYTAGVSPKKCVD 1390

Qy    420 VTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 449
          |||| :||||||||||||:   |||:|||||:
Db   1391 VTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422
```

```
RESULT 13
Q9F4JO
ID   Q9F4JO;
AC   Q9F4JO;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative outer membrane protein PG57.
OS   Porphyromonas gingivalis.
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA   Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT   "Identification of vaccine candidates from genomic analysis of
RT   Porphyromonas gingivalis.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY007522; AAG42428.1; -.
DR   InterPro; IPR003961; FN_III.
DR   SMART; SM00060; FN3; 1.
SQ   SEQUENCE   925 AA;  103632 MW;  5FF219BD6914DAE8 CRC64;

Query Match             28.0%;  Score 683;  DB 2;  Length 925;
Best Local Similarity   30.2%;  Pred. No. 1.3e-36;
Matches  176;  Conservative  62;  Mismatches 132;  Indels 212;  Gaps 22;

Qy     10 LSESFENG-IPASWKTIDADGDGNNWTTTPPGG-----TSFAGHNSAICASSASYINF 62
```

```
Db    311 LYENFENGPVPNGWLVIDADGDGFSW-----------GHYLNAYDAFPGHNGGHCSLSASYVPG 363
          |:|||||  | ||||||||||||| :|                              :|||||
```

```
Qy     63 EGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEE 122
          :|| ||||||||||:| |:||||   ||   :||   ||    :|||||||:||  :|:|
Db    364 IGPVTPDNYLITPKVE--GAKRVKYWVSTQDANWAAEHYAVMASTTGTAVGDFV-ILFEE 420
```

```
Qy    123 VLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV---- 177
          :|||    |:| ::                 :|||||||||:|:|||||||||||||
Db    421 TMTAKPT----------------------GAWYERTINLPEGTKYIAWRHYNCTDIYFLKLDDITVFGT ---- 467
```

```
Qy    178 ---------------------------------------------------E:KANG 183
                                                               |||||
Db    468 PASBPEPVTDFVVSLIENNKGRLKWNYPNGYEPDKTDDKDPLQLAGYNIYANGSLLVHIQ 527
```

```
Qy    184 ---------------KRADFTET--------------------------------FFSSTHGE 199
                         |||||||                              :||||||||
Db    528 DPTVLEYIDETYSSRDDQVEVEYCVTAVNDNIESQSVCDKLIYDSQSDIILYEGFEAGS 587
```

```
Qy    200 APAEWTTIDADGDGQGWLCLSSGQLDW-LTAHGGTNVVASFSWNGM-ALNPDNYLISKD 256
          |||||||||||||||||||||||||||||||||||||||::||||| ::
Db    588 IPEGWLLIDADGDGNVNW-----DYYPWTMYGHDSEKCIASPSYLPMIGVLTPDNYLVTPR 642
```

```
Qy    257 VTGATKVKYYYAVNDG-FPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTE 315
          :|:||||||||||||||:|:|||||||||||||||||||||||||:|:|||||||||| :||
Db    643 LEGAKLVKYWVSAQDAVVSAEHYAVMVSTTGTAVEDFVLLFEET---------------------MTAK 690
```

```
Qy    316 ADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDI----------------------- 357
          |:|:      ||:|||:||||||||||||||||||||:::||
Db    691 ANGA-----WYERTITLPAGTKYIAWRHYDCTDMFFLLLDDITVYRSTETVPEPVTDFVV 745
```

```
Qy    358 --------------------QFTMGGSPTPTD-------------YTYTVYRDGT------KIKEGLTETTF 390
                              :|||||||||||               |:||||||||       |||||||||
Db    746 SLIENNKGRLKWNYPNGYEPDKTDDKDPLQLTGYNIYANGSLLVHIQDPTVLEYIDETYS 805
```

```
Qy    391 EEDGVATGNHEYCVEVKYTAGV-SPKEC--VNVTV---DPVQ 426
          :|:|||| |:|:|:   |||||||  :||   :|:|   |||:
Db    806 SRDGQV--EMEYCVTAVNDNIESQSVCDKLNYTITSLDNIQ 845
```

```
RESULT 14
Q9KIB3
ID   Q9KIB3;                      PRELIMINARY;          PRT;    312 AA.
AC   Q9KIB3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical outer membrane protein PG27.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C.;
RT   "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF237555; AAF81413.1; -.
SQ   SEQUENCE   312 AA;  34592 MW;  0D5792C9643A25F5 CRC64;

Query Match             23.0%;  Score 561.5;  DB 2;  Length 312;
Best Local Similarity   32.8%;  Pred. No. 3.2e-29;
Matches  136;  Conservative  73;  Mismatches  23;  Indels 183;  Gaps  8;

Qy     10 LSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASYINFEGPQNPD 69
```

```
Db          |||||:||:||||||||||| ||            |||| ||  ::| |||||| ||||
          4 LSESFESGIPAIWKTIDADGDGYNWMHL-----TNFTG-QSGLCVSSASYIGGVGALTPD 57

Qy         ||||||     :: |||| |           ||||||:|::||:||:|  | ||| ||||
         70 NYLVTPELSLPNGG--TLTFWVCAQDANYASEHYAVIASSTGNDASNFANALLEEVLTAK 127

Db         ||:|||||  ::||| ||              |||||:|::|:|  ||| ||||||||||
         58 NYLITPELKLPTDALVEIIYWVCTQDLTAPSEHYAVYSSSTGNNAADFVNLLYEETLTAK 117

Qy        |||||||||||  |::||||| ||||||||||:|||||||||||||||||||||||||
        128 TVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGKRAD 187

Db         ::|:|||  ||:|||||||||||| |||||:||||||||:||:|:|
        118 R-IQSPELIRGNRTQGVWYQRKVVLPNDTKYVAFRHFNSTDNFWLNLLDEVSIL------ 169

Qy        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
        188 FTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASFSWNGMALN 247

Db
        170 ------------------------------------------------------------ 169

Qy        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        248 PDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGG 307

Db                                                           ||
        170 ----------------------------------------------YTP---------- 172

Qy        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
        308 ARFGLSTEADGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTP 367

Db                                                    ||
        173 -----------------------------LP--------RRAPC--------------PHP 182

Qy        |||||||||||||||||||| ||| :| :|||||||| ||||||||||||||
        368 TDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTV 422

Db         ||:|:|||| ||   ||  :| :: | ||||||||| :||:| ||||||||:|
        183 GGYTYSVFRDGQKIASGLSALAYIDTDVPYGTQDYCVQVNYLQGDSYKVCKNIVV 237
```

```
RESULT 15
Q9XBU9
ID   Q9XBU9       PRELIMINARY;      PRT;   293 AA.
AC   Q9XBU9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Immunoreactive 32 kDa antigen PG25.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF155351; AAD38410.1; -.
SQ   SEQUENCE   293 AA;  32272 MW;  CC03EAC241F7F6F1 CRC64;

Query Match             9.5%;   Score 233;  DB 2;  Length 293;
Best Local Similarity  59.0%;   Pred. No. 1.3e-07;
Matches   46;  Conservative   7;  Mismatches  23;  Indels   2;  Gaps   1;
```

```
Qy        ||||||||||||||||||||| ||| :| :|||||||| ||||||||||||||
        370 YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNP 429

Db         |||||||| ||   ||  :| :: | |||||||||  ::  ||||||||:|
         77 YTYTVYRDGVKIASGLTEPTFLDEDVPAGEHTYCVEVQYQGGVSDKVCVDVEVK--DFKP 134

Qy        |||||||||||||
        430 VQNLTGSAVGQKVTLKWD 447

Db         ||||| ::|:|| ||
        135 VTNLTGTASNDEVSLDWD 152
```

Search completed: May 18, 2004, 11:46:36
Job time : 34.8982 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:38:40 ; Search time 13.9034 Seconds
                                     (without alignments)
                                     1670.936 Million cell updates/sec

Title: US-08-570-311-16
Perfect score: 2443
Sequence: 1 FNPNPGTTTLSESFENGIPA.........QNLTGSAVGQKVTLKWDAPN 450

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_AA:*
             1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
             2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
             3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
             4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
             5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
             6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2443	100.0	450	2	US-08-570-311-16	Sequence 16, Appl
2	2443	100.0	2628	2	US-08-570-311-14	Sequence 14, Appl
3	2436	99.7	456	2	US-08-570-311-18	Sequence 18, Appl
4	2436	99.7	456	2	US-08-570-311-20	Sequence 20, Appl
5	2321	95.0	439	2	US-08-570-311-22	Sequence 22, Appl
6	2063	84.4	1477	4	US-09-482-500A-1	Sequence 1, Appli
7	2063	84.4	1687	2	US-08-570-311-29	Sequence 29, Appl
8	2063	84.4	1704	3	US-08-336-308A-10	Sequence 10, Appl
9	2063	84.4	1704	2	US-08-822-324-6	Sequence 6, Appli
10	2063	84.4	1704	4	US-09-490-931-10	Sequence 10, Appl
11	2051	84.0	1706	4	US-09-066-330-10	Sequence 10, Appl
12	2036	83.3	1087	2	US-08-570-311-8	Sequence 8, Appli
13	2036	83.3	1087	2	US-08-353-485-8	Sequence 8, Appli
14	2036	83.3	1358	2	US-08-570-311-27	Sequence 27, Appl
15	2016.5	82.5	1732	4	US-09-066-330-11	Sequence 11, Appl
16	2011.5	82.3	1732	2	US-08-570-311-10	Sequence 10, Appl
17	2011.5	82.3	1732	2	US-08-353-485-10	Sequence 10, Appl
18	546.5	22.4	497	2	US-08-570-311-2	Sequence 2, Appli
19	546.5	22.4	497	2	US-08-353-485-2	Sequence 2, Appli
20	296.5	12.1	942	1	US-08-141-124-14	Sequence 14, Appl
21	296.5	12.1	942	4	US-08-541-902-14	Sequence 14, Appl
22	236	9.7	942	3	US-08-822-324-18	Sequence 18, Appl
23	189	7.7	46	3	US-08-822-324-9	Sequence 9, Appli
24	135	5.5	1833	4	US-08-621-944A-4	Sequence 4, Appli
25	135	5.5	1833	4	US-08-945-567D-4	Sequence 4, Appli
26	135	5.5	1992	4	US-08-621-944A-3	Sequence 3, Appli
27	135	5.5	1992	4	US-08-945-567D-3	Sequence 3, Appli
28	133.5	5.5	2123	3	US-08-968-685A-10	Sequence 10, Appl
29	133.5	5.5	2736	4	US-09-252-991A-30227	Sequence 30227, A
30	130.5	5.3	492	4	US-09-48-500A-2	Sequence 2, Appli
31	130.5	5.3	737	3	US-08-119-361-5	Sequence 5, Appli
32	130.5	5.3	737	3	US-08-336-308A-4	Sequence 4, Appli
33	130.5	5.3	737	3	US-08-822-324-4	Sequence 4, Appli
34	130.5	5.3	737	4	US-09-490-931-4	Sequence 4, Appli
35	129	5.3	24	4	US-09-066-330-6	Sequence 6, Appli
36	128	5.2	25	3	US-08-822-324-19	Sequence 19, Appl
37	128	5.2	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
38	125.5	5.1	1095	4	US-08-851-567B-34	Sequence 34, Appl
39	124.5	5.1	2411	4	US-09-268-347-36	Sequence 36, Appl
40	122	5.0	509	4	US-08-822-324-8	Sequence 8, Appli
41	121.5	5.0	1638	4	US-09-071-035-258	Sequence 258, App
42	121.5	5.0	1638	4	US-09-071-035-262	Sequence 262, App
43	121.5	5.0	1638	4	US-09-071-035-266	Sequence 266, App
44	121.5	5.0	1747	4	US-09-134-000C-5999	Sequence 5999, Ap
45	121.5	5.0	2048	4	US-09-268-347-48	Sequence 48, Appl

ALIGNMENTS

RESULT 1
US-08-570-311-16

; Sequence 16, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

```
;          LENGTH: 450 amino acids
;          TYPE: amino acid
;          TOPOLOGY: linear
;          MOLECULE TYPE: protein
US-08-570-311-16

Query Match           100.0%;  Score 2443;  DB 2;  Length 450;
Best Local Similarity 100.0%;  Pred. No. 3.1e-203;
Matches 450;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 PNFNPGTTTLSESFENGIPASWKTIDADGDGNWWTTTPPGGTSFAGHNSAICASSASYI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 PNFNPGTTTLSESFENGIPASWKTIDADGDGNWWTTTPPGGTSFAGHNSAICASSASYI 60

Qy     61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120

Qy    121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKVVAFRHFGCTDFFWINLDDVEIK 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKVVAFRHFGCTDFFWINLDDVEIK 180

Qy    181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240

Qy    241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300

Qy    301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKVVAFRHYNCSDLNYILLDDIQFT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKVVAFRHYNCSDLNYILLDDIQFT 360

Qy    361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420

Qy    421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          ||||||||||||||||||||||||||||||
Db    421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450

RESULT 2
US-08-570-311-14
; Sequence 14, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;       APPLICANT: Progulske-Fox, Ann
;       APPLICANT: Tumwasorn, Somying
;       APPLICANT: Lepine, Guylaine
;       APPLICANT: Han, Naiming
;       APPLICANT: Lantz, Marilyn
;       APPLICANT: Patti, Joseph
;       TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;       TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;       NUMBER OF SEQUENCES: 29
;       CORRESPONDENCE ADDRESS:
;         ADDRESSEE: Ted W. Whitlock
;         STREET: 2421 N.W. 41st Street, Suite A-1
;         CITY: Gainesville
;         STATE: FL
;         COUNTRY: USA
;         ZIP: 32606
;       COMPUTER READABLE FORM:
;         MEDIUM TYPE: Floppy disk
;         COMPUTER: IBM PC compatible
;         OPERATING SYSTEM: PC-DOS/MS-DOS
;         SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;         APPLICATION NUMBER: US/08/570,311
;         FILING DATE:
;         CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;         APPLICATION NUMBER: US 08/353,485
;         FILING DATE: 09-DEC-1994
;         CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;         APPLICATION NUMBER: US 07/647,119
;         FILING DATE: 25-JAN-1991
;         CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;         APPLICATION NUMBER: US 07/241,640
;         FILING DATE: 08-SEP-1988
;       ATTORNEY/AGENT INFORMATION:
;         NAME: Whitlock, Ted W.
;         REGISTRATION NUMBER: 36,965
;         REFERENCE/DOCKET NUMBER: UF15.C3
;       TELECOMMUNICATION INFORMATION:
;         TELEPHONE: (904) 375-8100
;         TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 14:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH: 2628 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;         MOLECULE TYPE: protein
US-08-570-311-14

Query Match           100.0%;  Score 2443;  DB 2;  Length 2628;
Best Local Similarity 100.0%;  Pred. No. 4.5e-202;
Matches 450;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 PNFNPGTTTLSESFENGIPASWKTIDADGDGNWWTTTPPGGTSFAGHNSAICASSASYI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    500 PNFNPGTTTLSESFENGIPASWKTIDADGDGNWWTTTPPGGTSFAGHNSAICASSASYI 559

Qy     61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 619

Qy    121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKVVAFRHFGCTDFFWINLDDVEIK 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKVVAFRHFGCTDFFWINLDDVEIK 679

Qy    181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 739

Qy    241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 799

Qy    301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKVVAFRHYNCSDLNYILLDDIQFT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    800 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKVVAFRHYNCSDLNYILLDDIQFT 859

Qy    361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    860 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 919

Qy    421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          ||||||||||||||||||||||||||||||
Db    920 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 949

RESULT 3
US-08-570-311-18
; Sequence 18, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;       APPLICANT: Progulske-Fox, Ann
;       APPLICANT: Tumwasorn, Somying
;       APPLICANT: Lepine, Guylaine
;       APPLICANT: Han, Naiming
;       APPLICANT: Lantz, Marilyn
```

```
Qy   361 MGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 MGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 426

Qy   421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
             ||||||||||||||||||||||||||||||
Db   427 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

RESULT 4
US-08-570-311-20
; Sequence 20, Application US/08570311
; Patent No. 582791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 456 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-20

Query Match   99.7%; Score 2436; DB 2; Length 456;
Best Local Similarity   99.8%; Pred. No. 1.3e-202;
Matches 449; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

```
Qy     1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 66
```

Query Match   99.7%; Score 2436; DB 2; Length 456;
Best Local Similarity   99.8%; Pred. No. 1.3e-202;
Matches 449; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy     1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 66

Qy    61 NFEGPQONPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    67 NFEGPQONPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126

Qy   121 EEVLTAKTVVTAPEAIRGTRVQGTWYQGTKTVQLPAGIKYVAFRHFGCTDFFWINLDDVEIK 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127 EEVLTAKTVVTAPEAIRGTRVQGTWYQGTKTVQLPAGIKYVAFRHFGCTDFFWINLDDVEIK 186

Qy   181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFS 240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFS 246

Qy   241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306

Qy   301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366
```

```
;       REGISTRATION NUMBER: 36,965
;  REFERENCE/DOCKET NUMBER: UF15.C3
;  TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (904) 375-8100
;       TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO: 22:
;  SEQUENCE CHARACTERISTICS:
;       LENGTH: 439 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;  MOLECULE TYPE: protein
US-08-570-311-22

Query Match         95.0%;  Score 2321;  DB 2;  Length 439;
Best Local Similarity 98.4%;  Pred. No. 1.1e-192;
Matches 426;  Conservative   3;  Mismatches   4;  Indels   0;  Gaps   0;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASYI  60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||  |||||
Db   7 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICVSSASYI  66

Qy  61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
       |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  67 NFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126

Qy 121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db 127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIK 186

Qy 181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASFS 240
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASFS 246

Qy 241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306

Qy 301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366

Qy 361 MGGSPTPTDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db 367 MGGSPTPTDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNV 426

Qy 421 TVDPVQFNPVQNL 433
       |::|||||||||
Db 427 TINPTQFNPVQNL 439
```

RESULT 6
US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Imamura, Takahisa
; APPLICANT: Potempa, Jan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
; FILE REFERENCE: 235.00160101
; CURRENT APPLICATION NUMBER: US/09/482,500A
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/115,869
; PRIOR FILING DATE: 1999-01-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-482-500A-1

Query Match   84.4%; Score 2063; DB 4; Length 1477;

```
Qy  61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126

Qy 121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 186

Qy 181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASFS 240
       |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTNVVASFS 246

Qy 241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306

Qy 301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366

Qy 361 MGGSPTPTDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 367 MGGSPTPTDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 426

Qy 421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
       ||||||||||||||||||||||||||||||
Db 427 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

RESULT 5
US-08-570-311-22
; Sequence 22, Application US/08570311
; Patent No. 5824791
GENERAL INFORMATION:
APPLICANT: Progulske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.

```
Best Local Similarity 86.3%; Pred. No. 1.5e-169;
Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

Qy    1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-  59
          |||||||||||||||||||||||||||||||||||||||       ||:||::|||
Db  729 PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 785

Qy   60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          : |||||:|:|  |:|||:|||||||||||||||||||||||||||||||||||
Db  786 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 845

Qy  120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||| :|||| :||:||| :|||||||||||||||||| ||:|||||
Db  846 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 903

Qy  180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  904 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 963

Qy  240 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  964 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1023

Qy  300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1024 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1083

Qy  360 TMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1084 TMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1143

Qy  420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          |::| |||||||          |||||:|
Db 1144 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1174

RESULT 7
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Ted W. Whitlock
;     STREET: 2421 N.W. 41st Street, Suite A-1
;     CITY: Gainesville
;     STATE: FL
;     COUNTRY: USA
;     ZIP: 32606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/570,311
;     FILING DATE:
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/353,485
;     FILING DATE: 09-DEC-1994
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/647,119
;     FILING DATE: 25-JAN-1991
```

```
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/241,640
;     FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Whitlock, Ted W.
;     REGISTRATION NUMBER: 36,965
;     REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (904) 375-8100
;     TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 29:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1687 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-570-311-29

Query Match 84.4%; Score 2063; DB 2; Length 1687;
Best Local Similarity 86.3%; Pred. No. 1.8e-169;
Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

Qy    1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-  59
          |||||||||||||||||||||||||||||||||||||||       ||:||::|||
Db  939 PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 995

Qy   60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          : |||||:|:|  |:|||:|||||||||||||||||||||||||||||||||||
Db  996 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1055

Qy  120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||| :|||| :||:||| :|||||||||||||||||| ||:|||||
Db 1056 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1113

Qy  180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1114 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1173

Qy  240 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1174 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1233

Qy  300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1234 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1293

Qy  360 TMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1294 TMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1353

Qy  420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          |::| |||||||          |||||:|
Db 1354 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1384

RESULT 8
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
;   APPLICANT: Travis, James
;   APPLICANT: Potempa, Jan S.
;   APPLICANT: Barr, Philip J.
;   APPLICANT: Pavloff, Nadine
;   TITLE OF INVENTION: Porphyromonas gingivalis
;   TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
;     STREET: 5370 Manhattan Circle, Suite 201
;     CITY: Boulder
;     STATE: Colorado
```

GENERAL INFORMATION:
APPLICANT: Potempa, Jan S.
APPLICANT: Travis, James
APPLICANT: Genco, Caroline A.
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,324
FILING DATE: 21-MAR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/013,945
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 103-95 WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 488-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-822-324-6

Query Match   84.4%;   Score 2063;   DB 3;   Length 1704;
Best Local Similarity   86.3%;   Pred. No. 1.9e-169;
Matches 389;   Conservative 17;   Mismatches 39;   Indels 6;   Gaps 4;

```
Qy     1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTIPPPGGTSFAGHNSAICASSASY-  59
         ||||||||||||||||||||||||||||||||||            |  |||||||: :
Db   956 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1012

Qy    60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
         :  |  : |DNYL:TP:LDLPNGG||TFWVCAQDANYASEHYAVYASSTGNDASNF:NAL
Db  1013 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1072

Qy   120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
         |||:::AKG -VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI
Db  1073 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1130

Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1131 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1190

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1191 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1250

Qy   300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1251 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1310

Qy   360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1311 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1370
```

COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,308A
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/119,361
FILING DATE: 10-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,441
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 21-93C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-308A-10

Query Match   84.4%;   Score 2063;   DB 3;   Length 1704;
Best Local Similarity   86.3%;   Pred. No. 1.9e-169;
Matches 389;   Conservative 17;   Mismatches 39;   Indels 6;   Gaps 4;

```
Qy     1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTIPPPGGTSFAGHNSAICASSASY-  59
         ||||||||||||||||||||||||||||||||||            |  |||||||: :
Db   956 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1012

Qy    60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
         :  |  : |DNYL:TP:LDLPNGG||TFWVCAQDANYASEHYAVYASSTGNDASNF:NAL
Db  1013 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1072

Qy   120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
         |||:::AKG -VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI
Db  1073 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1130

Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1131 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 1190

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1191 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1250

Qy   300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1251 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1310

Qy   360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1311 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 1370

Qy   420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
         ||::: ||||| |:| |  :|| |:
Db  1371 VTINPTQFNPVXNLKAQPDGGDVVLKWEAPS 1401
```

RESULT 9
US-08-822-324-6
; Sequence 6, Application US/08822324
; Patent No. 6129917

```
Db    1131  KANGKRADFTETFESSTHGBAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASF  1190
                                                                                
Qy     240  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEET  299
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1191  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEET  1250

Qy     300  PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYHYNCSDLNYILLDDIQF  359
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1251  PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  1310

Qy     360  TMGGSPTPDYTVTVYRDGTKIKEGLIETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
             |||||||||||||||||||| ||| |||||||||||||||||||||||||||||||
Db    1311  TMGGSPTPDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  1370

Qy     420  VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  450
             ||:::|  |||:|:|||
Db    1371  VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS  1401
```

RESULT 11
US-09-066-330-10
; Sequence 10, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric C.
;   APPLICANT: Bhogal, Peter S.
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;   FILE REFERENCE: Reynolds
;   CURRENT APPLICATION NUMBER: US/09/066,330A
;   CURRENT FILING DATE: 1998-09-15
;   EARLIER APPLICATION NUMBER: PN 6275
;   EARLIER FILING DATE: 1995-10-30
;   EARLIER APPLICATION NUMBER: PCT/AU96/00673
;   EARLIER FILING DATE: 1996-10-30
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 1706
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-09-066-330-10

Query Match   84.0%; Score 2051; DB 4; Length 1706;
Best Local Similarity   85.8%; Pred. No. 2e-168;
Matches 387; Conservative 18; Mismatches 40; Indels   6; Gaps   4;

```
Qy      1  PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-  59
            |||||||||||||||||||||||||||||||||||||| ::|                |:
Db    958  PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG  1014

Qy     60  INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
            :   |||||||:||| |||:|||||||||||||||||||||||||||||||||||||
Db   1015  LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL  1074

Qy    120  LEEVLTAKTVVTAPEAIRGTRVQGTWVQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
            |::|  ||| |::|| |||  ||||||||||:|||||||||||||||||| |:|:||:|
Db   1075  LEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI  1132

Qy    180  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASF  239
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVSSF  1192

Qy    240  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEET  299
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEET  1252

Qy    300  PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYHYNCSDLNYILLDDIQF  359
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1253  PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  1312

Qy    360  TMGGSPTPDYTVTVYRDGTKIKEGLIETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN  419
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

(left column, rotated)

```
Qy    420  VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  450
            ||:::|  |||:|:|||
Db   1371  VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS  1401
```

RESULT 10
US-09-490-931-10
; Sequence 10, Application US/09490931
; Patent No. 6274718
; GENERAL INFORMATION:
;   APPLICANT: Travis, James
;   APPLICANT: Potempa, Jan S.
;   APPLICANT: Barr, Philip J.
;   APPLICANT: Pavloff, Nadine
;   TITLE OF INVENTION: Porphyromonas gingivalis
;   TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
;   STREET: 5370 Manhattan Circle, Suite 201
;   CITY: Boulder
;   STATE: Colorado
;   COUNTRY: US
;   ZIP: 80303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/490,931
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/336,308
;   FILING DATE:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/265,441
;   FILING DATE: 24-JUN-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ferber, Donna M.
;   REGISTRATION NUMBER: 33,878
;   REFERENCE/DOCKET NUMBER: 21-93C
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (303) 499-8080
;   TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1704 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-490-931-10

Query Match   84.4%; Score 2063; DB 3; Length 1704;
Best Local Similarity   86.3%; Pred. No. 1.9e-169;
Matches 389; Conservative 17; Mismatches 39; Indels   6; Gaps   4;
```

```
Qy      1  PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY-  59
            |||||||||||||||||||||||||||||||||||||| ::|                |:
Db    956  PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG  1012

Qy     60  INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
            :   |||||||:||| |||:|||||||||||||||||||||||||||||||||||||
Db   1013  LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL  1072

Qy    120  LEEVLTAKTVVTAPEAIRGTRVQGTWVQKTVQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
            |::|  ||| |::|| |||  ||||||||||:|||||||||||||||||| |:|:||:|
Db   1073  LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI  1130

Qy    180  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASF  239
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db   396 LGGIGIVLTPDNYLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 455

Qy   120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ||:|:||:|| ::|:|| :|:: ||:||||||||||||||||| : ::|:||
Db   456 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 513

Qy   180 KANGKRADFTETFESSTHGEAPAEWTTIDADGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   514 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGTNVVASF 573

Qy   240 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   574 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 633

Qy   300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
          |||||||||||||||||||||||||||||||||||||||||||||||:||||||:|||
Db   634 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYILLDDIQF 693

Qy   360 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   694 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVN 753

Qy   420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          ||:::|||||:||||:| | ||| :||
Db   754 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 784

RESULT 13
US-08-353-485-8
; Sequence 8, Application US/08353485
; Patent No. 5830710
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:

Db   1313 TMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVN 1372

Qy   420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
           ||:::|||||:||||:| | ||| :||
Db   1373 VTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 12
US-08-570-311-8
; Sequence 8, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1087 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-8

Query Match   83.3%; Score 2036; DB 2; Length 1087;
Best Local Similarity 85.4%; Pred. No. 2e-167;
Matches 385; Conservative 18; Mismatches 42; Indels 6; Gaps 4;

Qy   1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFPAGHNSAICASSASY- 59
          ||||||||||||||||||||||||||||||||| |:||
Db   339 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAFG---IAGYNSNGCVYSESFG 395

Qy   60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          : | ||||||||||||||||||||||||||||||||||||||||||||||

```
;    LENGTH: 1087 amino acids
;    TYPE: amino acid
;    TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-353-485-B

Query Match           83.3%;   Score 2036;  DB 2;  Length 1087;
Best Local Similarity  85.4%;   Pred. No. 2e-167;
Matches 385;  Conservative  18;  Mismatches 42;  Indels 6;  Gaps 4;

Qy     1  PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY- 59
          ||||||||||||||||||||||||||||||||||| :| |||||||  :|| ||||||||
Db   339  PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 395

Qy    60  INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          ||||||||:  |||||  ||||||||||||||||||||||||||||||||||||||||||
Db   396  LGGIGVLTPDNYLITPALDLANGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 455

Qy   120  LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          |||||||| ||| ||:||||:|||||||||||||||||||||||||| ||| :|:||||
Db   456  LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 513

Qy   180  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   514  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 573

Qy   240  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   574  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 633

Qy   300  PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   634  PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYILLDDIQF 693

Qy   360  TMGGSPTPTDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   694  TMGGSPTPTDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVN 753

Qy   420  VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          || :|||||||  |||||:  |||:| |||:|
Db   754  VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 784
```

```
Query Match           83.3%;   Score 2036;  DB 2;  Length 1358;
Best Local Similarity  85.4%;   Pred. No. 2.9e-167;
Matches 385;  Conservative  18;  Mismatches 42;  Indels 6;  Gaps 4;

Qy     1  PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASY- 59
          ||||||||||||||||||||||||||||||||||| :| |||||||  :|| ||||||||
Db   610  PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 666

Qy    60  INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          ||||||||:  |||||  ||||||||||||||||||||||||||||||||||||||||||
Db   667  LGGIGVLTPDNYLITPALDLANGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 726

Qy   120  LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          |||||||| ||| ||:||||:|||||||||||||||||||||||||| ||| :|:||||
Db   727  LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 784

Qy   180  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   785  KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF 844

Qy   240  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   845  SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 904

Qy   300  PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   905  PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYILLDDIQF 964

Qy   360  TMGGSPTPTDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   965  TMGGSPTPTDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVN 1024

Qy   420  VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          || :|||||||  |||||:  |||:| |||:|
Db  1025  VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1055
```

```
;    CURRENT APPLICATION NUMBER: US/09/066,330A
;    CURRENT FILING DATE: 1998-09-15
;    EARLIER APPLICATION NUMBER: PN 6275
;    EARLIER FILING DATE: 1995-10-30
;    EARLIER APPLICATION NUMBER: PCT/AU96/00673
;    EARLIER FILING DATE: 1996-10-30
;    NUMBER OF SEQ ID NOS: 15
;    SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;    LENGTH: 1732
;    TYPE: PRT
;    ORGANISM: Porphyromonas gingivalis
US-09-066-330-11

Query Match            82.5%;   Score 2016.5;  DB 4;   Length 1732;
Best Local Similarity  85.0%;   Pred. No. 2e-165;
Matches   384;  Conservative   18;  Mismatches   41;  Indels    9;  Gaps    6;

Qy      1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASY-     59
          |||||| ||||||||||||||||||||||||                 |:|| |  :|:
Db    977 PNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1032

Qy     60 INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
          :     | || |||| || || ||||||||||||||||||||||||||||||   |||
Db   1033 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1092

Qy    120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
          |||:||  :|   |  ||:||  : || |||||||||||||||||||  :|||:||:||:
Db   1093 LEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1150

Qy    180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASF  239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||:|:||
Db   1151 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGSNVVSSF 1210

Qy    240 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET  299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 SWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1270

Qy    300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF  359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1330

Qy    360 TMGGSPTPTDYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPKECVN    419
          |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db   1331 TMGGSPTPTDYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPKKCVN  1390

Qy    420 VTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP  449
          ||: ||:|:|:||| |  |  |:|||||::||
Db   1391 VTVNSTQFNFVQNLTAEQAPNSMDAILKWNAP 1422
```

Search completed: May 18, 2004, 11:49:13
Job time : 14.9034 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May 18, 2004, 11:42:50 ; Search time 34.8564 Seconds
                                 (without alignments)
                            3592.387 Million cell updates/sec

Title:           US-08-570-311-16
Perfect score:   2443
Sequence:        1 PNPNPGTTTLSESFENGIPA.........QNLTGSAVGQKVTLKWDAPN 450

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters:  ,    1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published Applications AA:*
            1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
            2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
            3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
            4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
            5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
            6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
            7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
            8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
            9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
           10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
           11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
           12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
           13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
           14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
           15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
           16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
           17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
           18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result          Query
No.     Score   Match   Length  DB  ID                          Description
------  ------  ------  ------  --  ---------------------       -----------
  1     2051    84.0    1706    14  US-10-229-066-10            Sequence 10, Appl
  2     2016.5  82.5    1732    14  US-10-229-066-11            Sequence 11, Appl
  3     700     28.7    419     15  US-10-174-695-3             Sequence 3, Appli
  4     680.5   27.9    419     15  US-10-174-695-5             Sequence 6, Appli
  5     385     15.8    231     15  US-10-174-695-6             Sequence 4, Appli
  6     206.5   8.5     196     15  US-10-174-695-4             Sequence 18, Appl
  7     153     6.3     29      15  US-10-387-977-18            Sequence 21, Appl
  8     153     6.3     29      15  US-10-387-977-21            Sequence 23, Appl
  9     153     6.3     29      15  US-10-387-977-23            Sequence 24, Appl
 10     153     6.3     29      15  US-10-387-977-24            Sequence 26, Appl
 11     153     6.3     29      15  US-10-387-977-26            Sequence 55467, A
 12     148     6.1     872     12  US-10-282-122A-55467        Sequence 17, Appl
 13     147     6.0     29      15  US-10-387-977-17            Sequence 25, Appl
 14     145     5.9     29      15  US-10-387-977-25            Sequence 101, App
 15     144     5.9     509     15  US-10-387-977-101
```

```
 16     136     5.6     29      15  US-10-387-977-20            Sequence 20, Appl
 17     135     5.5     1833    14  US-10-175-282-4             Sequence 4, Appli
 18     135     5.5     1833    14  US-10-175-275-4             Sequence 4, Appli
 19     135     5.5     1992    14  US-10-175-282-3             Sequence 3, Appli
 20     135     5.5     1992    14  US-10-175-275-3             Sequence 3, Appli
 21     134     5.5     25      15  US-10-387-977-77            Sequence 77, Appl
 22     134     5.5     25      15  US-10-387-977-80            Sequence 80, Appl
 23     134     5.5     27      15  US-10-387-977-11            Sequence 11, Appl
 24     134     5.5     27      15  US-10-387-977-12            Sequence 12, Appl
 25     134     5.5     27      15  US-10-387-977-13            Sequence 13, Appl
 26     133.5   5.5     2122    9   US-09-813-214A-9            Sequence 9, Appli
 27     130.5   5.3     2468    12  US-10-282-122A-66335        Sequence 66335, A
 28     130.5   5.3     2468    14  US-10-246-330-4             Sequence 4, Appli
 29     130     5.3     25      15  US-10-387-977-81            Sequence 81, Appl
 30     129     5.3     24      14  US-10-229-066-6             Sequence 6, Appli
 31     129     5.3     2435    12  US-10-282-122A-47453        Sequence 47453, A
 32     128     5.2     1946    12  US-10-282-122A-62947        Sequence 62947, A
 33     127     5.2     25      15  US-10-387-977-83            Sequence 83, Appl
 34     127     5.2     25      15  US-10-387-977-84            Sequence 84, Appl
 35     127     5.2     25      15  US-10-387-977-85            Sequence 85, Appl
 36     126     5.2     1649    15  US-10-369-493-18460         Sequence 18460, A
 37     125.5   5.1     555     14  US-10-156-761-14592         Sequence 14592, A
 38     125.5   5.1     1095    14  US-10-242-056-34            Sequence 34, Appl
 39     125.5   5.1     1095    15  US-10-262-794A-34           Sequence 34, Appl
 40     124.5   5.1     691     15  US-10-369-493-23473         Sequence 23473, A
 41     124.5   5.1     1333    12  US-10-282-122A-55546        Sequence 5546, A
 42     123     5.0     31      15  US-10-387-977-6             Sequence 6, Appli
 43     122     5.0     491     15  US-10-387-977-100           Sequence 100, App
 44     121.5   5.0     1638    12  US-10-206-576-258           Sequence 258, App
 45     121.5   5.0     1638    12  US-10-206-576-262           Sequence 262, App
```

ALIGNMENTS

```
RESULT 1
US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US20030157637A1
US-10-229-066-10
; GENERAL INFORMATION:
;  APPLICANT: Reynolds, Eric C.
;  APPLICANT: Bhogal, Peter S.
;  APPLICANT: Slakeski, Nada
;  TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;  FILE REFERENCE: Reynolds
;  CURRENT APPLICATION NUMBER: US/10/229,066
;  CURRENT FILING DATE: 2002-08-28
;  PRIOR APPLICATION NUMBER: US/09/066,330
;  PRIOR FILING DATE: 1998-09-15
;  PRIOR APPLICATION NUMBER: PN 6275
;  PRIOR FILING DATE: 1995-10-30
;  PRIOR APPLICATION NUMBER: PCT/AU96/00673
;  PRIOR FILING DATE: 1996-10-30
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 10
;  LENGTH: 1706
;  TYPE: PRT
;  ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match             84.0%;  Score 2051;  DB 14;  Length 1706;
Best Local Similarity   85.8%;  Pred. No. 9.8e-176;
Matches  387;  Conservative  18;  Mismatches  40;  Indels  6;  Gaps  4;

Qy     1  PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASSASY-  59
          ||||||||||||||||||||||||||||||||||||||||||||||||| ::||: ::
Db   958  PNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG  1014

Qy    60  INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
          :||||| :|||||:|||:|||:|||:||||||||||||||||||||||||||||||||||
Db  1015  LGGIGVLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL  1074
```

```
Qy    120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ::|:|||| |::|||| :|  |||||:|||||:||||||||||||| |:||:|:|||||
Db   1075 LEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1132

Qy    180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASF 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db   1133 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVSSF 1192

Qy    240 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1252

Qy    300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1253 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1312

Qy    360 TMGGSPPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db   1313 TMGGSPPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVN 1372

Qy    420 VTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
          |||: :||||||||  :| |  :||||:||
Db   1373 VTVNSTQFNPVQNLKAQPDGGDVVLKWEAPS 1403
```

RESULT 2
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

```
Query Match      82.5%;  Score 2016.5;  DB 14;  Length 1732;
Best Local Similarity   85.0%;  Pred. No. 1.3e-172;
Matches  384;  Conservative  18;  Mismatches  41;  Indels  9;  Gaps  6;

Qy      1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSPAGHNSAICASSASY- 59
          ||||||:||||||||||||||||||||||||  :   |||||| ||||  |  :||
Db    977 PNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 1032

Qy     60 INFEGPQNPDNYLVTPELSLPNGGLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          :    | :|||||:||| ||||||:||||||||||||||||||||||||||||||:|||
Db   1033 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 1092

Qy    120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ::|:|||| |::|||| :|  |||||:|||||:||||||||||||| |:||:|:|||||
Db   1093 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 1150

Qy    180 KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASF 239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||::|:||
Db   1151 KANGKRADFTETFESSTHGEAPABWTTIDADGDGQWLCLSSGQLDWLTAHGGSNVVSSF 1210

Qy    240 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 SWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEET 1270

Qy    300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 359
          ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db   1271 PNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQF 1330

Qy    360 TMGGSPPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVN 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db   1331 TMGGSPPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVN 1390

Qy    420 VTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 449
          |||: :||||||||     :    : ||:|||
Db   1391 VTVNSTQFNPVQNLTABQAPNSMDAILKWNAP 1422
```

RESULT 3
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

```
Query Match      28.7%;  Score 700;  DB 15;  Length 419;
Best Local Similarity   74.7%;  Pred. No. 1.8e-54;
Matches  139;  Conservative  11;  Mismatches  30;  Indels  6;  Gaps  4;

Qy      1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSPAGHNSAICASSASY- 59
          ||||||||||||||||||||||||||||||  :   |||||| ||||  |  :||
Db    239 PNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG 295

Qy     60 INFEGPQNPDNYLVTPELSLPNGGLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
          :    | :|||||:||| ||||||:||||||||||||||||||||||||||||||:|||
Db    296 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL 355

Qy    120 LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI 179
          ::|:|||| |::|||| :|  |||||:|||||:||||||||||||| |:||:|:|||||
Db    356 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 413

Qy    180 KANGKR 185
          ||||||
Db    414 KANGKR 419
```

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928000700
; CURRENT APPLICATION NUMBER: US/10/174,695

```
;   CURRENT FILING DATE: 2002-06-18
;   PRIOR APPLICATION NUMBER: PCT/AU00/01588
;   PRIOR FILING DATE: 2000-12-21
;   PRIOR APPLICATION NUMBER: AU PQ 4859
;   PRIOR FILING DATE: 1999-12-24
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-174-695-5

  Query Match             27.9%;     Score 680.5;  DB 15;  Length 419;
  Best Local Similarity   73.7%;     Pred. No. 1e-52;
  Matches   137;  Conservative  12;  Mismatches  30;  Indels  7;  Gaps  5;

   Qy     1  PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASY-   59
             ||||||  |||||||||||||||||||||||||||||||| |  :  |||   :  :
   Db   240  PNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYSESFG  295

   Qy    60  INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL  119
             : |  ::|: : ||||||||||||||||||||||||||||||||||||||||||||
   Db   296  LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNAL  355

   Qy   120  LEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEI  179
             |: :|||   :  :||| : |||||||:||||||||||||||| :|:|:  :|:|
   Db   356  LEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI  413

   Qy   180  KANGKR  185
             ||||||
   Db   414  KANGKR  419

RESULT 5
US-10-174-695-6
; Sequence 6, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: Slakeski, Nada
;   APPLICANT: Chen, Chao Guang
;   APPLICANT: Barr, Ian George
;   TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
;   FILE REFERENCE: 52928000700
;   CURRENT APPLICATION NUMBER: US/10/174,695
;   CURRENT FILING DATE: 2002-06-18
;   PRIOR APPLICATION NUMBER: PCT/AU00/01588
;   PRIOR FILING DATE: 2000-12-21
;   PRIOR APPLICATION NUMBER: AU PQ 4859
;   PRIOR FILING DATE: 1999-12-24
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 231
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-174-695-6

  Query Match             15.8%;     Score 385;  DB 15;  Length 231;
  Best Local Similarity   41.4%;     Pred. No. 2.2e-26;
  Matches   103;  Conservative  31;  Mismatches  69;  Indels  46;  Gaps  11;

   Qy   207  IDADGDGQGWLCLSSGQLDWLTAHGGTNVV-ASFSWNGMALNPDNYLISKD--VTGATKV  263
             |::::|| |: ::|| |  :  :::: |  :   :|:  : ::|
   Db     3  LDADHNTFGSVIPATGPL--FTGTASSNLYSANFEYL-IPANADPVTTQNIIVTGQGEV   59

   Qy   264  KYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEADGAKPQS  323
             |:                   :
   Db    60  V------IPGGVY---------DYCITNPEASG------KMWIAGDGGNQPA   91

   Qy   324  VWIERTVDLPAGTKYV-AFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIK  382
             ::  : |: ||  |||   :  ||||||||
   Db    92  RYDDFTFE--AGKKYTFTMRRAGMGDGTDMEVED-------DSPASYTYTVYRDGTKIK  141

   Qy   383  EGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVD-PVQPNPVQNLTGSAVGQK  441
             |||||  ||||||||||||||||||||||||||||||||||   : ||||::||||||
   Db   142  EGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCVCKDVTVEGSNEFAPVQNLTGSSVGQK  201

   Qy   442  VTLKWDAPN  450
             |||||||||
   Db   202  VTLKWDAPN  210

RESULT 6
US-10-174-695-4
; Sequence 4, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: Slakeski, Nada
;   APPLICANT: Chen, Chao Guang
;   APPLICANT: Barr, Ian George
;   TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
;   FILE REFERENCE: 52928000700
;   CURRENT APPLICATION NUMBER: US/10/174,695
;   CURRENT FILING DATE: 2002-06-18
;   PRIOR APPLICATION NUMBER: PCT/AU00/01588
;   PRIOR FILING DATE: 2000-12-21
;   PRIOR APPLICATION NUMBER: AU PQ 4859
;   PRIOR FILING DATE: 1999-12-24
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 196
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-174-695-4

  Query Match              8.5%;     Score 206.5;  DB 15;  Length 196;
  Best Local Similarity   37.5%;     Pred. No. 2.4e-10;
  Matches    57;  Conservative  13;  Mismatches  63;  Indels  19;  Gaps  5;

   Qy     1  PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI   60
             ||||||  |||||||||||||||||||||||||||||  : |||:||   :|||||||||
   Db    48  PNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYLDNSA  104

   Qy    61  NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL  120
             :  ::   ||||||||||||||||||||||||||||||
   Db   105  KIDRNQBINVY-NTAEYAKTNNAPI------KVVGYADE------KTCTAAYNM--KLS  148

   Qy   121  EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQL  152
             :|: ||||||||||||||||||| :|:
   Db   149  ERRAKAVAKMLEKYGVSADRITIEWKGSSEQI  180

RESULT 7
US-10-387-977-18
; Sequence 18, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928000301
;   CURRENT APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
```

```
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 18
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-18

Query Match          6.3%;   Score 153;   DB 15;   Length 29;
Best Local Similarity  100.0%;   Pred. No. 1e-06;
Matches    29;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy    369  DYTYTVYRDGTKIKEGLTETTFEEDGVAT 397
                |||||||||||||||||||||||||||||
Db      1  DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 8
US-10-387-977-21
;   Sequence 21, Application US/10387977
;   Publication No. US20040005276A1
;   GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928200301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 21
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-21

Query Match          6.3%;   Score 153;   DB 15;   Length 29;
Best Local Similarity  100.0%;   Pred. No. 1e-06;
Matches    29;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy    369  DYTYTVYRDGTKIKEGLTETTFEEDGVAT 397
                |||||||||||||||||||||||||||||
Db      1  DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 9
US-10-387-977-23
;   Sequence 23, Application US/10387977
;   Publication No. US20040005276A1
;   GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928200301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
```

```
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 23
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-23

Query Match          6.3%;   Score 153;   DB 15;   Length 29;
Best Local Similarity  100.0%;   Pred. No. 1e-06;
Matches    29;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy    369  DYTYTVYRDGTKIKEGLTETTFEEDGVAT 397
                |||||||||||||||||||||||||||||
Db      1  DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 10
US-10-387-977-24
;   Sequence 24, Application US/10387977
;   Publication No. US20040005276A1
;   GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928200301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 24
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-24

Query Match          6.3%;   Score 153;   DB 15;   Length 29;
Best Local Similarity  100.0%;   Pred. No. 1e-06;
Matches    29;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy    369  DYTYTVYRDGTKIKEGLTETTFEEDGVAT 397
                |||||||||||||||||||||||||||||
Db      1  DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 11
US-10-387-977-26
;   Sequence 26, Application US/10387977
;   Publication No. US20040005276A1
;   GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928200301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
```

```
Qy    4 NPGTTLSESFENGIPASWKTIDADGDGNWTTPPGGTSFAGHNSAICASSASYINFE  63
          |||:|::|:.:||::|:.:|      :|:.|:.:.|::|:
Db  293 NTGTGTHNVTVDTGLPS--VAFNAISDDNVLNAVEKGQDLSVSGTSANLAEGTVVTVTLN 350

Qy   64 GPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASST--GNDASNFANALL  120
          |:  |  |:  : :|  ||::||...| .| |: .:|.
Db  351 G---KNYAATTAAD---GTWSLTVPAADLAGLGDHYTLSATATNGVGNSVSNTANLLV  402

Qy  121 EEVLFAKT------VVTAPEAIRGTRVQGTWYQKTVQLPAG--TKYVAFRHFGCT-  167
           :|      :  .:| ...| .:    ::
Db  403 DTALPTVTINTVAGDNVINAAEVAAGQTISG---KVANAEAGNTVTVTIIGGNSYTATVQ 458

Qy  168 -DFFW-INLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWICLSSGQLD  225
          . :|.:  : .   :| .: :. .|:           ::
Db  459 SDLTWSVNVPESVLTALGNGDLTVSATVTNGHGNTGAGEREIVIDAS------LPGLRID 512

Qy  226 WL----TAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHY  278
          :      : :: ::|  |
Db  513 TVAGDDVINSIEHGQNLIVTGSSDGLA-------------AGTTLT--VTVN---GKTY 553

Qy  279 AVMISKTGT----------NAGDFTVFEETPNGINKGGARFGLSTEADGAKPQSVW  325
          | |:   :|          :|
Db  554 AASVLADGTWSAAIPAADVGALAAGTVTV---TAAGQSAAG---------NPVTIS  597

Qy  326 IERTVDLPAGTKYVAFRHYNCSDLNYILLDDI-----------QPTMGGSPTPTD-  369
          : |||||:.|   :|    : :|    ::||:
Db  598 HDVTVDLSA----VAI----SIDAIATDDVINAAEKGADLVLSGGSTSNVEENQTVTIT 647

Qy  370 -----YTYTVVRDG------TKIKEG---LTETTFEEDGVATGNHEYCVEVKYT  409
           :|:.| ||:   |       ||   .|    .  :| .|
Db  648 SGGKTYTAKVDADGNWTATVPSADLLAGLKDGDASVQVSVTNAHGNSASAGREYSVD---- 703

Qy  410 AGVSPKECVNVTVDPVQFNPVQNLTGSA------------VGQKVTLKWD  447
          |.  .:  .||:.|  .|.|      :|            :|||.|:|
Db  704 -ATAP----TVTIDTVAGDNVINASEAAAGVAISGTTTAEVGQTVTVTLD 748
```

```
Query Match    6.0%; Score 147; DB 15; Length 29;
Best Local Similarity 96.6%; Pred. No. 3.6e-06;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  369 DYTYTVVRDGTKIKEGLTETTFEEDGVAT 397
          ||||||||||||||||||  ||||||||
Db    1 DYTYTVVRDGTKIKEGLTATTFEEDGVAT  29
```

```
Query Match    6.3%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  369 DYTYTVVRDGTKIKEGLTETTFEEDGVAT 397
          |||||||||||||||||||||||||||||
Db    1 DYTYTVVRDGTKIKEGLTETTFEEDGVAT  29
```

```
Query Match    6.1%; Score 148; DB 12; Length 872;
Best Local Similarity 21.7%; Pred. No. 0.00041;
Matches 115; Conservative 59; Mismatches 196; Indels 160; Gaps 26;
```

```
Qy    106  SSTGNDASNFANALL---EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQ--LPAGTKYVA 160
           :||.|:|  ||  :.|:  :|:.||::|    ||  ||.||||.||  .|||| :||
Db    123  ASSPEELTNIIDKVLMVEKATMPDKSYLEKVLLIAG--ADYSWNSQVGQPTIKYGMQYYY 180

Qy    161  FRHFGCTDFFWINLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL--C 218
           .:  |||:     :     :.  |.:|                     :. :
Db    181  NQEHGYTDVY---------NYLKAPYTGCYSHLNTGVSFANYT---AHGSETAWADPL 226

Qy    219  LSSGQLDWLT------AHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVND 271
           |::.|:||   ||     :.|:      :|:     :|
Db    227  LTTSQLKALTNKDKYFLAIGNCCITAQFDY----VQP----CFGEVITRVKEKGAYAYIG 278

Qy    272  GFPGDH----------YAVMISKTGTNAGDFTVVF-EETPNGINKGGARFGLSTEA 316
           .:  |:.   || :|  |:.|.:  ||:.|:
Db    279  SSPNSYWGEDYTWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVN---------- 326

Qy    317  DGAKPQSVWIERTVDLPAGT---KYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT- 371
           ||  :|       :||  :  .: :.||
Db    327  -----SIMWAGNLAATHAGNIGNITHIGAHYY--WEAYHVLGDGSVMPYRAMPKTNTYTL 379

Qy    372  --------------YTVYRDGTKIKEGLTETTFEEDGVAT---------GNHE 401
           :.||       |||:   :|.|        |.::
Db    380  PASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANAS----GVATVSMTKQITENGNYD 435

Qy    402  YCVEVKYTAGVSPKECVNTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
           :.     .|| :||.||.:||      .   |.| :.|||||.|:.
Db    436  VVITRSNYLPVIKQIQVG---EPSPYQPVSNLTATTQGQKVTLKWEAPS 481


Search completed: May 18, 2004, 11:52:20
Job time : 35.8564 secs
```

RESULT 14
US-10-387-977-25
; Sequence 25, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200003301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-25

Query Match   5.9%; Score 145; DB 15; Length 29;
Best Local Similarity   96.6%; Pred. No. 5.4e-06;
Matches   28; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   369 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 397
               |||||||||||||||.|||||||||||||
Db   1 DYTYTVYRDGTKIKWGLTETTFEEDGVAT 29

RESULT 15
US-10-387-977-101
; Sequence 101, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200003301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-101

Query Match   5.9%; Score 144; DB 15; Length 509;
Best Local Similarity   20.0%; Pred. No. 0.00043;
Matches   94; Conservative   66; Mismatches 189; Indels 120; Gaps   21;

Qy   48 HNSAICASSAS-YINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAV-YA 105
           | :| |.|. | |:|.| :|| |: ||.|.:| : |.| |.:   .||| ||
Db   67 YNDGLAASAAPVFLAIVG----DTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMS 122

```
                 GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 18, 2004, 11:33:39 ; Search time 47.0287 Seconds
                                     (without alignments)
                                     2739.638 Million cell updates/sec

Title:      US-08-570-311-18
Perfect score:  2480
Sequence:   1 GTPNPNPNPGTTTLSESF.........QNLTGSAVGQKVTLKWDAPN 456

Scoring table:   BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:   158610 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:   1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  A_Geneseq.29Jan04:*
            1: geneseqp1980s:*
            2: geneseqp1990s:*
            3: geneseqp2000s:*
            4: geneseqp2001s:*
            5: geneseqp2002s:*
            6: geneseqp2002as:*
            7: geneseqp2003bs:*
            8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2480	100.0	456	2	AAR96023	Aar96023 P. gingiv
2	2480	100.0	456	2	AAR96022	Aar96022 P. gingiv
3	2480	100.0	456	2	AAW69491	Aaw69491 Haemagglu
4	2480	100.0	456	2	AAW69490	Aaw69490 Haemagglu
5	2480	100.0	2628	2	AAR96030	Aar96030 P. gingiv
6	2480	100.0	2628	2	AAW69488	Aaw69488 Haemagglu
7	2436	98.2	450	2	AAR96021	Aar96021 P. gingiv
8	2436	98.2	450	2	AAW69489	Aaw69489 Haemagglu
9	2351	94.8	439	2	AAR96024	Aar96024 P. gingiv
10	2351	94.8	439	2	AAW69492	Aaw69492 Haemagglu
11	2082	84.0	1687	2	AAR96033	Aar96033 P. gingiv
12	2082	84.0	1687	2	AAW69495	Aaw69495 Haemagglu
13	2082	84.0	1704	2	AAR70188	Aar70188 Arg-gingi
14	2082	84.0	1704	3	AAW34843	Aaw34843 Arg-gingi
15	2082	84.0	1704	3	AAY67396	Aay67396 Arg-gingi
16	2082	84.0	1704	4	AAU08938	Aau08938 P. gingiv
17	2070	83.5	1706	2	AAW24786	Aaw24786 PrtR anti
18	2066	83.3	1087	2	AAR96028	Aar96028 P. gingiv
19	2066	83.3	1087	2	AAW69486	Aaw69486 Haemagglu
20	2066	83.3	1358	2	AAR96032	Aar96032 P. gingiv
21	2066	83.3	1358	2	AAW69494	Aaw69494 Haemagglu
22	2034.5	82.0	1732	2	AAR96029	Aar96029 P. gingiv
23	2034.5	82.0	1732	2	AAW24787	Aaw24787 PrtK anti
24	2034.5	82.0	1732	2	AAW69487	Aaw69487 Haemagglu
25	726	29.3	419	4	AAU03572	Aau03572 P. gingiv
26	720	29.0	135	6	ABP55081	Abp55081 Porphyrom
27	710.5	28.6	419	4	AAU03574	Aau03574 P. gingiv
28	704	28.4	134	4	AAB49217	Aab49217 Peptide u
29	682	27.5	921	2	AAY34522	Aay34522 Porphyrom
30	682	27.5	922	2	AAY34521	Aay34521 Porphyrom
31	682	27.5	925	2	AAY34520	Aay34520 Porphyrom
32	682	27.5	938	2	AAY34392	Aay34392 Porphyrym
33	565.5	22.8	377	2	AAY34359	Aay34359 Porphyrym
34	561.5	22.6	312	2	AAY34484	Aay34484 Porphyrym
35	546.5	22.0	497	2	AAR96025	Aar96025 P. gingiv
36	546.5	22.0	497	2	AAW69483	Aaw69483 Haemagglu
37	415	16.7	970	2	AAR72458	Aar72458 Porphyrom
38	385	15.5	231	4	AAU03575	Aau03575 P. gingiv
39	267	10.8	991	2	AAR77313	Aar77313 Porphyrom
40	236	9.5	49	2	AAW34805	Aaw34805 Arg-speci
41	233	9.4	293	2	AAY34483	Aay34483 Porphyrom
42	233	9.4	299	2	AAY34358	Aay34358 Porphyrom
43	232.5	9.4	196	4	AAU03573	Aau03573 P. gingiv
44	189	7.6	46	2	AAW34798	Aaw34798 Arg-speci
45	153	6.2	29	2	AAW83085	Aaw83085 Peptide f

ALIGNMENTS

```
RESULT 1
AAR96023
ID   AAR96023 standard; protein; 456 AA.
XX
AC   AAR96023;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep3 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep3.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A,  Tumwasorn S,  Lepine G,  Han N,  Lantz M,  Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30647.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 110-112; 153pp; English.
XX
CC   HArep3 (AAR96023) is the product of the HArep3 repeat unit (AAT30647) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep3 and other hagA repeat
CC   unit products (see also AAR96021-22 and AAR96024) can be obtd. from
CC   transformed host cells and used as vaccines to protect humans or animals
CC   against periodontal disease. Expression in Salmonella cells allows prodn.
CC   of live vaccine. HArep-4 can also be used to detect the presence of anti
CC   -P. gingivalis antibodies and to raise monoclonal antibodies for
CC   diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 456 AA;
```

```
CC  HArep2 (AAR96022) is the product of the HArep2 repeat unit (AAT30646) of
CC  the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC  haemagglutinin hagA (see also AAR96030). HArep2 and other hagA repeat
CC  unit products (see also AAR96021 and AAR96023-24) can be obtd. from
CC  transformed host cells and used as vaccines to protect humans or animals
CC  against periodontal disease. Expression in Salmonella cells allows prodn.
CC  of live vaccine. HArep1-4 can also be used to detect the presence of anti
CC  -P. gingivalis antibodies and to raise monoclonal antibodies for
CC  diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ  Sequence 456 AA;

Query Match             100.0%;  Score 2480;  DB 2;  Length 456;
Best Local Similarity   100.0%;  Pred. No. 7.5e-191;
Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456

RESULT 3
AAW69491
ID   AAW69491 standard; protein; 456 AA.
XX
AC   AAW69491;
XX
DT   22-DEC-1998  (first entry)
XX
DE   Haemagglutinin protein hagA, HArep3.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
```

```
Query Match             100.0%;  Score 2480;  DB 2;  Length 456;
Best Local Similarity   100.0%;  Pred. No. 7.5e-191;
Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456

RESULT 2
AAR96022
ID   AAR96022 standard; protein; 456 AA.
XX
AC   AAR96022;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep2 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep2.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Proguiske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30646.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 107-108; 153pp; English.
XX
```

PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58877.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 127-132; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 456 AA;

Query Match   100.0%; Score 2480; DB 2; Length 456;
Best Local Similarity   100.0%; Pred. No. 7.5e-191;
Matches 456; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy   421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db   421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456

RESULT 5
AAR96030
ID   AAR96030 standard; protein; 2628 AA.
XX

PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58878.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 133-138; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 456 AA;

Query Match   100.0%; Score 2480; DB 2; Length 456;
Best Local Similarity   100.0%; Pred. No. 7.5e-191;
Matches 456; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy   421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db   421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456

RESULT 4
AAW69490
ID   AAW69490 standard; protein; 456 AA.
XX
AC   AAW69490;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagA, HArep2.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX

```
AC    AAR96030;
XX
DT    16-OCT-2003 (revised)
DT    04-SEP-1996 (first entry)
XX
DE    P. gingivalis haemagglutinin hagA.
XX
KW    Haemagglutinin; hagA; periodontal disease; vaccine; antibody.
XX
OS    Porphyromonas gingivalis; strain 381.
XX
FH    Key             Location/Qualifiers
FT    Peptide         5..21
FT                    /label= Sig_peptide
XX
PN    WO9617936-A2.
XX
PD    13-JUN-1996.
XX
PF    11-DEC-1995;  95WO-US016108.
XX
PR    09-DEC-1994;  94US-00353485.
XX
PA    (UYFL ) UNIV FLORIDA.
PA    (UABR-) UAB RES FOUND.
XX
PI    Progulske-Fox A,  Tunwasorn S,  Lepine G,  Han N,  Lantz M,  Patti JM;
XX
DR    WPI; 1996-287181/29.
DR    N-PSDB; AAT30654.
XX
PT    Porphyromonas gingivalis genes and proteins - used in the detection and
PT    vaccination against periodontal disease.
XX
PS    Claim 6; Page 93-101; 153pp; English.
XX
CC    P. gingivalis 381 haemagglutinin hagA (AAR96030) was identified as the
CC    product of the hagA gene (AAT30654) isolated as an EcoRV fragment of
CC    genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
CC    24), can be obtd. from transformed host cells and used as a vaccine to
CC    protect humans or animals against periodontal disease. Expression in
CC    Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
CC    also be used to detect the presence of anti-P. gingivalis antibodies and
CC    to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-
CC    2003 to standardise OS field)
XX
SQ    Sequence 2628 AA;

Query Match             100.0%;  Score 2480;  DB 2;  Length 2628;
Best Local Similarity   100.0%;  Pred. No. 9e-190;
Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    950 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 1009

Qy     61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1069

Qy    121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 1129

Qy    181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 1189

Qy    241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTV 1249

Qy    301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1250 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1309

Qy    361 DDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 DDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1369

Qy    421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db   1370 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 1405


RESULT 6
AAW69488
ID    AAW69488 standard; protein; 2628 AA.
XX
AC    AAW69488;
XX
DT    22-DEC-1998  (first entry)
XX
DE    Haemagglutinin protein hagA.
XX
KW    Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS    Porphyromonas gingivalis.
XX
PN    US5824791-A.
XX
PD    20-OCT-1998.
XX
PF    11-DEC-1995;    95US-00570311.
XX
PR    08-SEP-1988;    88US-00241640.
PR    25-JAN-1991;    91US-00647119.
PR    09-DEC-1994;    94US-00353485.
XX
PA    (UYFL ) UNIV FLORIDA.
PA    (UABR-) UAB RES FOUND.
XX
PI    Patti JM,  Han N,  Lantz M,  Tunwasorn S,  Progulske-Fox A,  Lepine G;
XX
DR    WPI; 1998-582627/49.
DR    N-PSDB; AAV58875.
XX
PT    Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT    protease poly:peptide(s)).
XX
PS    Claim 1; Col 91-110; 101pp; English.
XX
CC    This sequence is encoded by a Porphyromonas gingivalis gene of the
CC    invention. This sequence represents the hagA haemagglutinin protein. The
CC    polypeptides are used to produce antibodies to organisms associated with
CC    periodontal disease. The antibodies are also used in purification and
CC    identification procedures. The genes and polypeptides are used as
CC    vaccines against periodontal disease
XX
SQ    Sequence 2628 AA;

Query Match             100.0%;  Score 2480;  DB 2;  Length 2628;
Best Local Similarity   100.0%;  Pred. No. 9e-190;
Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    950 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 1009

Qy     61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1069

Qy    121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 1129
```

```
Qy    181  DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130  DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 1189

Qy    241  VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190  VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 1249

Qy    301  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1309

Qy    361  DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310  DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1369

Qy    421  KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
                ||||||||||||||||||||||||||||||||||||
Db   1370  KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 1405
```

RESULT 7
AAR96021
ID   AAR96021 standard; protein; 450 AA.
XX
AC   AAR96021;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArepl product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArepl.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995; 95WO-US016108.
XX
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30645.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 103-104; 153pp; English.
XX
CC   HArepl (AAR96021) is the product of the HArepl repeat unit (AAT30645) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArepl and other hagA repeat
CC   unit products (see also AAR96022-24) can be obtd. from transformed host
CC   cells and used as vaccines to protect humans or animals against
CC   periodontal disease. Expression in Salmonella cells allows prodn. of live
CC   vaccine. HArepl-4 can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 450 AA;

Query Match   98.2%; Score 2436; DB 2; Length 450;
Best Local Similarity 99.8%; Pred. No. 2.6e-187;
Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      7  PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 66
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60

Qy     67  NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120

Qy    127  EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 186
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180

Qy    187  ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGTNVVASFS 246
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGTNVVASFS 240

Qy    247  WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300

Qy    307  NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360

Qy    367  MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 426
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420

Qy    427  TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
               |||||||||||||||||||||||||||||
Db    421  TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450
```

RESULT 8
AAW69489
ID   AAW69489 standard; protein; 450 AA.
XX
AC   AAW69489;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagA, HArep1.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995; 95US-00570311.
XX
PR   08-SEP-1988; 88US-00241640.
PR   25-JAN-1991; 91US-00647119.
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58876.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 121-126; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The

```
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 450 AA;

Query Match            98.2%;  Score 2436;  DB 2;  Length 450;
Best Local Similarity  99.8%;  Pred. No. 2.6e-187;
Matches  449;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    7 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI   66
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI   60

Qy   67 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL  126
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL  120

Qy  127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK  186
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK  180

Qy  187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTNVVASFS   246
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASFS   240

Qy  247 WNGMALNPDNYLIISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP  306
         ||||||||||||||||||||||||||||||||||||||||||||
Db  241 WNGMALNPDNYLIISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP  300

Qy  307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT  366
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT  360

Qy  367 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV  426
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV  420

Qy  427 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  456
         ||||||||||||||||||||||||||||
Db  421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  450

RESULT 9
AAR96024
ID   AAR96024 standard; protein; 439 AA.
XX
AC   AAR96024;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep4 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep4.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;  95WO-US016108.
PR   09-DEC-1994;  94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
```

```
DR   N-PSDB; AAT30648.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 114-115; 153pp; English.
XX
CC   HArep4 (AAR96024) is the product of the HArep4 repeat unit (AAT30648) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep4 and other hagA repeat
CC   unit products (see also AAR96021-23) can be obtd. from transformed host
CC   cells and used as vaccines to protect humans or animals against
CC   periodontal disease. Expression in Salmonella cells allows prodn. of live
CC   vaccine. HArep1-4 can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 439 AA;

Query Match            94.8%;  Score 2351;  DB 2;  Length 439;
Best Local Similarity  98.2%;  Pred. No. 1.7e-180;
Matches  431;  Conservative   3;  Mismatches   5;  Indels   0;  Gaps   0;

Qy    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV   60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN  120
         ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN  120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL  180

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTN   240
         | ||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db  181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTN   240

Qy  241 VVASFSWNGMALNPDNYLIISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLIISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV  300

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL  360

Qy  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP  420

Qy  421 KECVNVTVDPVQFNPVQNL  439
         | |||| :: ||||||||
Db  421 KVCVNVTINPTQFNPVQNL  439

RESULT 10
AAW69492
ID   AAW69492 standard; protein; 439 AA.
XX
AC   AAW69492;
XX
DT   22-DEC-1998  (first entry)
XX
DE   Haemagglutinin protein hagA, HArep4.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
```

```
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58879.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 139-144; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 439 AA;

Query Match             94.8%;  Score 2351;  DB 2;  Length 439;
Best Local Similarity   98.2%;  Pred. No. 1.7e-180;
Matches   431;  Conservative   3;  Mismatches   5;  Indels   0;  Gaps   0;

Qy      1 GTPNPNPNGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICA  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTPNPNPNGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICv  60

Qy     61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy    121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy    181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240

Qy    241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy    301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy    361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy    421 KECVNVTVDPVQFNPVQNL 439
          |||   |||:|||||||||
Db    421 KVCVNVTINPTQFNPVQNL 439


RESULT 11
AAR96033
ID   AAR96033 standard; protein; 1687 AA.
XX
AC   AAR96033;
XX
DT   16-OCT-2003  (revised)
```
```
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagE.
XX
KW   Haemagglutinin; hagE; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain FDC381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30656.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 138-143; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagE (AAR96033) was identified as the
CC   product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
CC   The haemagglutinin can be obtd. from transformed host cells and used as a
CC   vaccine to protect humans or animals against periodontal disease.
CC   Expression in Salmonella cells allows prodn. of a live vaccine. The
CC   haemagglutinin can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1687 AA;

Query Match             84.0%;  Score 2082;  DB 2;  Length 1687;
Best Local Similarity   86.2%;  Pred. No. 5.2e-158;
Matches   392;  Conservative   17;  Mismatches   40;  Indels   6;  Gaps   4;

Qy      3 PNPNPNPNGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASS  62
          ||||||||||||||||||||||||||||||||||||||||||||||||||     |:|:
Db    935 PNPNPNPNGTTTLSESFENGIPASWKTIDADGDGHGHGWKPGNAPG---IAGYNSNGCVVS  991

Qy     63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          ||: ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    992 ESFGLGGIGVLTPDNYLJTPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy    122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          |||||||||||||||::|  |  |:|:  ::|||||||||||||||||||||||  :|:|
Db   1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTMFYIDLD 1109

Qy    182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
          :|||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   1110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1169

Qy    242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1229

Qy    302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1289

Qy    362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1349

Qy    422 ECVNVTVDPVQFNPVQNLTGSSAVGQKVTLKWDAPN 456
```

```
Db   1350 ECVNVTINPTQFNPVKNLKAQPDGDVVLKWEAPS 1384
          |||||||::|| |||||:||     |||||:||

RESULT 12
AAW69495
ID   AAW69495 standard; protein; 1687 AA.
XX
AC   AAW69495;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagE.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagE.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995; 95US-00570311.
XX
PR   08-SEP-1988;  88US-00241640.
PR   25-JAN-1991;  91US-00647119.
PR   09-DEC-1994;  94US-00053485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tunwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58881.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 167-182; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagE haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1687 AA;

Query Match    84.0%;    Score 2082;  DB 2;   Length 1687;
Best Local Similarity   86.2%;   Pred. No. 5.2e-158;
Matches  392;   Conservative   17;   Mismatches   40;   Indels    6;   Gaps    4;

Qy      3 PNPNPNENPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASS 62
          |||||||||||||||||||||||||||||||      ||:|: ||| ||
Db    935 PNPNPNENPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 991

Qy     63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          :|:            |||||  :||||||  ||||||||||||||||||||||||||||
Db    992 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy    122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          |||||  |   |  |:   ||:| |  | |::||:|||||||||||||| :|:::||
Db   1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy    182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSQGLWLTAHGGTNV 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSQGLWLTAHGGTNV 1169

Qy    242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1229
```

```
Qy    302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1289

Qy    362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 1349

Qy    422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          |||||:|      |||:|:| ||||||||:|||:|
Db   1350 ECVNVTINPTQFNPVKNLKAQPDGDVVLKWEAPS 1384

RESULT 13
AAR70188
ID   AAR70188 standard; protein; 1704 AA.
XX
AC   AAR70188;
XX
DT   25-MAR-2003 (revised)
DT   21-SEP-1995 (first entry)
XX
DE   Arg-gingipain-2 prepolyprotein.
XX
KW   Arg-gingipain-2; gingivalis; periodontal disease; vaccine;
KW   arginine-specific protease.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key            Location/Qualifiers
FT   Protein        228..719
FT                  /label= Protease
FT                  /note= "corresponds to Arg-gingipain-1"
FT   Region         720..1091
FT                  /label= Hemagglutinin
FT   Region         1092..1429
FT                  /label= Hemagglutinin
FT   Region         1430..1704
FT                  /label= Hemagglutinin
XX
PN   WO9507286-A1.
XX
PD   16-MAR-1995.
XX
PF   09-SEP-1994;  94WO-US010283.
XX
PR   10-SEP-1993;  93US-00119361.
PR   21-OCT-1993;  93US-00141324.
PR   24-JUN-1994;  94US-00265441.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J, Potempa J, Barr PJ, Pavloff N;
XX
DR   WPI; 1995-123373/16.
DR   N-PSDB; AAQ83489.
XX
PT   DNA encoding Arg-gingipain proteins - used to develop prods. for
PT   detection, treatment and prevention of periodontal disease.
XX
PS   Disclosure; Page 70-77; 89pp; English.
XX
CC   A low mol.wt. arginine-specific gingipain (AG-1) and high mol.wt. AG (AG-
CC   2) were isolated from P. gingivalis strains H66 (ATCC 33277) and W50
CC   (ATCC 53973). The sequences of the proteins were used to design PCR
CC   primers and probes to isolate AG DNA. Lambda DASH and lambda ZAP
CC   libraries were screened with a probe based on amino acids 11-22 of the AG
CC   protein to obtain DNA encoding AG-1 (AAQ83484) and AG-2 (AAQ83489). AG-2
CC   is a prepolyprotein incorporating AG-1. (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 1704 AA;
```

Query Match
Best Local Similarity   84.0%;   Score 2082;   DB 2;   Length 1704;
                        86.2%;   Pred. No. 5.2e-158;
Matches 392;   Conservative 17;   Mismatches 40;   Indels 6;   Gaps 4;

```
Qy    3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNWWTTTPPGGTSFAGHNSAICASS 62
        ||||||||||||||||||||||||||||||||||||       :|:|| |
Db  952 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
         || : ||||||||||||||||||||||||||||||||||||||||
Db 1009 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
        ||||| :||| |:|||||||| ||||||||:||||||||||||||| :|:||
Db 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy  182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
        :|||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db 1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy  242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy  302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
        |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db 1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy  362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1307 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy  422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
        |||||: :||: ||||:|:   |:||
Db 1367 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

RESULT 14
AAW34843
ID   AAW34843 standard; protein; 1704 AA.
XX
AC   AAW34843;
XX
DT   03-JUN-1998 (first entry)
XX
DE   Arg-gingipain high molecular weight prepolyprotein sequence.
XX
KW   Arg-specific gingipain protease; gingivalis; periodontal disease;
KW   vaccine; infection.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key   Location/Qualifiers
FT   Protein   1..227
FT   /note= "precursor protein"
XX
PN   WO9734629-A1.
XX
PD   25-SEP-1997.
XX
PF   21-MAR-1997; 97WO-US004635.
XX
PR   22-MAR-1996; 96US-0013945P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
PA   (MORE-) MOREHOUSE SCHOOL MEDICINE.
XX
PI   Potempa J, Travis J, Genco C;
XX
DR   WPI; 1997-479993/44.
DR   N-PSDB; AAT93872.
XX

PT   Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
PT   useful for protecting animals and humans from gingivalis and periodontal
PT   diseases.
XX
PS   Disclosure; Page 68-73; 95pp; English.
XX
CC   The present sequence represents an arginine-specific protease of
CC   Porphyromonas gingivalis. The following peptides, derived from Arg- and
CC   Lys-specific high molecular weight proteases, offer protection against
CC   infection: YTYTVRDGK IKEGLTATTE DDGVATGNHE YCVEKYTAGS VSPKVC (I);
CC   YTPVEEKQNG RMVIVAKKY (II); QLPFIFDVAC VNGDFLFSMP CFAEALMRAQ (III);
CC   GERPNYQPVS NLTATTQGQK VTLKWDAPSTK (IV); GNHEYCVEVK YTAGVSPKVC KDVTV (V);
CC   RMFMNYEPGR YTPVEEKQNG (VI); TFAGFEDTYK RMPMNYEPGR (VII); DYTYTVYRDG
CC   TKIKEGLTAT TFEEDGVATG NMEYCVCHKY TAGVSPKVC (VIII); YTYTVYRDGT KIKEGLTATTF
CC   EEDG (IX); RDGTKIKEGL TATTFEEDGV ATGN (X); KIKEGLTATT FEEDGVATGN HEY (XI)
CC   ; KWDAPNGTPN PNPN PNPNPN PGTTLSE (XII); and YTPVEEKENG RMIVIVAKKY
CC   (XIII). They are used in vaccines to protect animals, including humans,
CC   from gingivitis and/or periodontal diseases
XX
SQ   Sequence 1704 AA;

Query Match
Best Local Similarity   84.0%;   Score 2082;   DB 2;   Length 1704;
                        86.2%;   Pred. No. 5.2e-158;
Matches 392;   Conservative 17;   Mismatches 40;   Indels 6;   Gaps 4;

```
Qy    3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNWWTTTPPGGTSFAGHNSAICASS 62
        ||||||||||||||||||||||||||||||||||||       :|:|| |
Db  952 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
         || : ||||||||||||||||||||||||||||||||||||||||
Db 1009 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
        ||||| :||| |:|||||||| ||||||||:||||||||||||||| :|:||
Db 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy  182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
        :|||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db 1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy  242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy  302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
        |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db 1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy  362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1307 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy  422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
        |||||: :||: ||||:|:   |:||
Db 1367 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

RESULT 15
AAY67396
ID   AAY67396 standard; protein; 1704 AA.
XX
AC   AAY67396;
XX
DT   25-APR-2000 (first entry)
XX
DE   Arg-gingipain-2 amino acid sequence.
XX
KW   Arginine specific proteinase; Arg-gingipain; gingipain; gingipain-2;
KW   immunogenic component; vaccine; inflammatory response; tissue damage;
KW   periodontal disease.
XX
OS   Porphyromonas gingivalis.

```
XX
FH   Key              Location/Qualifiers
FT   Region           229..719
FT                    /note= "Amino acids 229-719 are specifically claimed"
FT   Region           720..1185
FT                    /note= "Amino acids 720-1185 are specifically claimed"
XX
PN   US6017532-A.
XX
PD   25-JAN-2000.
XX
PF   08-NOV-1994;     94US-00336308.
XX
PR   10-SEP-1993;     93US-00119361.
PR   24-JUN-1994;     94US-00265441.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Potempa JS, Travis J;
XX
DR   WPI; 2000-136659/12.
DR   N-PSDB; AAZ60181.
XX
PT   New Porphyromonas gingivalis arginine-specific protease preparation
PT   useful for preparing vaccines against periodontal disease and for
PT   screening for Arg-gingipain inhibitors.
XX
PS   Claim 1; Col 29-42; 55pp; English.
XX
CC   This sequence represents a Porphyromonas gingivalis arginine-specific
CC   proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
CC   Gingipain-2 consists or a 50kD protease component non-covalently
CC   associated with a 44kD haemagglutinin component. The proteinase is
CC   stimulated by glycine containing peptides and glycine analogues. It is
CC   inhibited by cysteine protease group specific inhibitors. The protease
CC   preparation can be used in immunogenic compositions and vaccines against
CC   inflammatory response and tissue damage caused by P. gingivalis in
CC   periodontal disease. It can also be used to screen for agents that
CC   modulate Arg-gingipain proteinase activity inhibitors
XX
SQ   Sequence 1704 AA;

Query Match             84.0%;  Score 2082;  DB 3;  Length 1704;
Best Local Similarity   86.2%;  Pred. No. 5.2e-158;
Matches  392;  Conservative  17;  Mismatches  40;  Indels  6;  Gaps  4;

Qy      3  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASS   62
           |||||||||||||||||||||||||||||||||||||||||||  :  :||:||| :|
Db    952  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGMWKPGNAPG---IAGYNSNGCVYS 1008

Qy     63  ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF  121
           | :|  |||||: :: |: ||||| |||||||||||||||||||||||||||||||||||
Db   1009  ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy    122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD  181
           ||:|||:|::|||| |:| ||| | ||| ||||||||||||||||||||| : |:|||:|
Db   1069  TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy    182  DVEIKANGKRADPTETFESSTHGEAPAEWTTIDADGDQGWLCLSSGQLGWLTAHGGTNV  241
           :|||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db   1127  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDQGWLCLSSGQLDWLTAHGGTNV 1186

Qy    242  VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGRPGDHYAVMLSKTGTNAGDFTVV  301
           :||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db   1187  VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFTVV 1246

Qy    302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1247  FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy    362  DIQFTMGGSPFTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307  DIQFTMGGSPFTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy    422  ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  456
           |||||||::|:||:|: |||||||||||||||||
Db   1367  ECVNVTINFTQFNPVKNLKAQPDGGDVVLKWEAPS 1401

Search completed: May 18, 2004, 11:42:42
Job time : 49.0287 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:37:00 ; Search time 11.906 Seconds
(without alignments)
3684.135 Million cell updates/sec

Title: US-08-570-311-18
Perfect score: 2480
Sequence: 1 GTPNPNPNPGTTLSESF.........QNLTGSAVGQKVTLKWDAPN 456

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2480	100.0	2628	2	T28651	hemagglutinin A -
2	2082	84.0	1704	2	A55426	gingipain R (EC 3.
3	2059	83.0	1526	2	S49763	gingipain R (EC 3.
4	2034.5	82.0	1732	2	T30836	lysine-specific cy
5	266	10.7	991	2	I40229	arginyl endopeptid
6	139.5	5.6	1052	2	AF2959	conserved hypothet
7	139.5	5.6	1341	2	H98323	hypothetical prote
8	132.5	5.3	691	2	B75622	hypothetical prote
9	131.5	5.3	2468	2	A83412	hypothetical prote
10	128.5	5.2	1248	2	C89874	autolysin [importe
11	128	5.2	4936	2	AH2515	hypothetical prote
12	127.5	5.1	1684	2	S10789	amylase A-180 - al
13	127	5.1	713	2	B75489	hypothetical prote
14	126	5.1	1441	2	A86685	prophage pil prote
15	126	5.1	1649	2	C86822	hypothetical prote
16	124.5	5.0	1904	2	T13256	tail-host specific
17	124.5	5.0	5291	2	F90696	hypothetical prote
18	123.5	5.0	5188	2	B85547	probable RTX famil
19	122.5	4.9	1274	2	T10729	transferrin-like p
20	121.5	4.9	465	2	A47023	s-layer protein -
21	121.5	4.9	3624	2	AD0835	large repetitive p
22	121	4.9	1090	2	S59077	cellulose 1,4-beta
23	120	4.8	908	2	AE2254	hypothetical prote
24	120	4.8	2783	2	T34416	hypothetical prote
25	119	4.8	715	2	JC4908	alkaline serine pr
26	119	4.8	1873	2	T30944	surface protein pr
27	118.5	4.8	926	2	D86897	hypothetical prote
28	118.5	4.8	1385	2	T18213	parasporal crystal
29	118.5	4.8	2817	2	B97033	uncharacterized pr
30	118.5	4.8	4199	2	S76412	hypothetical prote
31	118	4.8	1461	2	E90696	hypothetical prote
32	118	4.8	1461	2	A85547	hypothetical prote
33	117.5	4.7	1034	2	T30551	beta-galactosidase
34	116.5	4.7	1635	2	AI0452	hemolysin [importe
35	115.5	4.7	702	2	S48753	major surface prot
36	115.5	4.7	875	2	AF0472	probable outer mem
37	115.5	4.7	1118	2	A49724	protein-tyrosine-p
38	115.5	4.7	1282	2	JC4393	microbial collagen
39	115.5	4.7	1939	2	D97316	probable s-layer p
40	115	4.6	821	2	AD1507	probable secreted
41	115	4.6	1651	2	JC1340	outer membrane pro
42	115	4.6	3029	2	S76109	hypothetical prote
43	114	4.6	872	2	S49541	cellulae - Cellul
44	113	4.6	1345	2	H90975	hypothetical prote
45	112.5	4.5	938	2	AF1772	internalin-like pr

ALIGNMENTS

RESULT 1
T28651
hemagglutinin A - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C;Accession: T28651
R;Han, N.; Whitlock, J.; Progulske-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A;Title: The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381 contains four 1
A;Reference number: Z20494; MUID:97047672; PMID:8926061
A;Accession: T28651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2628 <HAN>
A;Cross-references: EMBL:U41807; NID:g1552410; PID:g1469916; PIDN:AAB17128.1
C;Genetics:
A;Gene: hagA

Query Match 100.0%; Score 2480; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 1.4e-162;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   950 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 1009

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1069

Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAPRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAPRHFGCTDFFWINL 1129

Qy   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 1189

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 1249

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1309

Qy   361 DDIQFTMGGSPTPDTYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 DDIQFTMGGSPTPDTYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1369

Qy   421 KECVNVTVDVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          |||||||||||||||||||||||||||||||||||

Db   1370 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 1405

RESULT 2
A55426
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
N;Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP; R
C;Species: Porphyromonas gingivalis
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
C;Accession: A55426; D53113
R;Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr, E
J. Biol. Chem. 270, 1007-1010, 1995
A;Title: Molecular cloning and structural characterization of the Arg-gingipain proteina
A;Reference number: A55426; MUID:95138080; PMID:7836351
A;Accession: A55426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1704 <PAV>
A;Cross-references: GB:U15282; NID:g557067; PIDN:AAA69539.1; PID:g557068
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: D53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 228-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141694)
C;Keywords: cysteine proteinase; hydrolase

Query Match   84.0%;   Score 2082;   DB 2;   Length 1704;
Best Local Similarity   86.2%;   Pred. No. 2.2e-135;
Matches 392; Conservative   17; Mismatches   40; Indels   6; Gaps   4;

Qy   3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
           |||||||||||||||||||||||||||||||||||||| :| :|| ||||||||
Db   952 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1009 ESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
           ||||||||||||||||||||||||||||||||||||||||||||||||:| |||:||||
Db   1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTNV 241
           :||||||||||||||||||||||||||||||||||||||||||||||:|| ||||||||
Db   1127 EVEIKANGKRADFTETFESSTHGEATAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPDTYTYVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 421
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307 DIQFTMGGSPTPDTYTYVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy   422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
           :||||||||||||:| :|||
Db   1367 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401

RESULT 3
S49763
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997

C;Accession: S49763
R;Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.
submitted to the EMBL Data Library, November 1994
A;Description: Cloning, sequence analysis and expression in Escherichia coli of prpR1 of
A;Reference number: S49763
A;Accession: S49763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <ADU>
A;Cross-references: EMBL:X82680
C;Genetics:
A;Gene: prpR1
C;Keywords: cysteine proteinase; hydrolase

Query Match   83.0%;   Score 2059;   DB 2;   Length 1526;
Best Local Similarity   85.5%;   Pred. No. 7.4e-134;
Matches 389; Conservative   17; Mismatches   43; Indels   6; Gaps   4;

Qy   3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
           |||||||||||||||||||||||||||||||||||| :| :|| ||||||||
Db   954 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011 ESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

Qy   122 ANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
           ||||||||||||||||||||||||||||||||||||||||||||||||:| |||:||||
Db   1071 TNALLEETITAKG-VRSPEAIRG-RIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTNV 241
           :||||||||||||||||||||||||||||||||||||||||||||||:|| ||||||||
Db   1129 EVEIKANGKRADFTETFESSTHGEATAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNV 1188

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db   1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1249 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy   362 DIQFTMGGSPTPDTYTYVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 421
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1309 DIQFTMGGSPTPDTYTYVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy   422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
           :||||||||||||:| :|||
Db   1369 KCVNVTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 4
T30836
lysine-specific cysteine proteinase porphypain (EC 3.4.22.-) - Porphyromonas gingivalis
N;Alternate names: lysine-specific cysteine proteinase 1, 60K
C;Species: Porphyromonas gingivalis
C;Date: 22-Oct-1999 #sequence revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30836; T30837; T30526; A53113
R;Barkocy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulske-Fox, A.; Lantz
J. Bacteriol. 178, 2734-2741, 1996
A;Title: Analysis of the prtP gene encoding porphypain, a cysteine proteinase of Porphy
A;Reference number: Z20895; MUID:96213011; PMID:8631659
A;Accession: T30836
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1732 <BAR>
A;Cross-references: EMBL:U42210; NID:g1314325; PIDN:g1314326; PID:AAB06565.1
R;Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z20896
A;Accession: T30837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA

A;Residues: 1-795,'T',797-1389,'N',1391-1478,'Y',1480-1732 <SLA>
A;Cross-references: EMBL:U75366; NID:g2182811; PID:g2182812; PIDN:AAB60809.1
R;Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A;Title: IS195, an insertion sequence-like element associated with protease genes in Por
A;Reference number: Z20844; MUID:98298016; PMID:9632563
A;Accession: T30526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1350,'N',1352-1363,'Y',1365-1447,'H',1449-1732 <LEW>
A;Cross-references: EMBL:AF017059; NID:g2738802; PID:g2738803; PIDN:AAC26523.1
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: A53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 229-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141690)
C;Genetics:
A;Gene: prtP; prtK
C;Keywords: cysteine proteinase; hydrolase

Query Match　　82.0%;　　Score 2034.5;　DB 2;　Length 1732;
Best Local Similarity　84.3%;　　Pred. No. 4.3e-132;
Matches　387;　Conservative　19;　Mismatches　44;　Indels　9;　Gaps　6;

Qy　　　1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAIC 59
　　　　　　|||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db　　969 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy　　 60 ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
　　　　　　|||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 1026 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy　　119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
　　　　　　|||| |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db　 1086 SNFTNALLEEFITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy　　179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHG 238
　　　　　　|||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy　　239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
　　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy　　299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
　　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy　　359 LLDDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
　　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 1324 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy　　419 SPKBCVNVTVDPVQFNPVQNLTGSAV--GQKVTLFKWDAP 455
　　　　　　|| :||| ||::||||||:||||||:　:||||||||||
Db　 1384 SPKKCVDVIVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422

RESULT 5
I40229
arginyl endopeptidase - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40229
R;Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A;Title: Structural characterization of arginpain, a novel arginine-specific cysteine
A;Reference number: I40229; MUID:95168884; PMID:7864651
A;Accession: I40229

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-991 <RES>
A;Cross-references: GB:D26470; NID:g927644; PIDN:BAA05484.1; PID:g927645

Query Match　　10.7%;　　Score 266; DB 2; Length 991;
Best Local Similarity　25.9%;　　Pred. No. 2e-10;
Matches　114;　Conservative　50;　Mismatches 164;　Indels 112;　Gaps 19;

Qy　　 30 TIDADGDGNNWTTTPPGGTSFAGHNSAICASSASYINFEGPQNPDNYLVTPELSLPNGG 89
　　　　　　 :::　:|::::||||||||||　||::|　:::::::|||| :|:
Db　　595 SVNVSCDYNGAIATISANGKMF---GSAAVENGTATINLTG-----------LTNES 637

Qy　　 90 TLTFWVCAQDANYASEHYAVYASSTG--NDASNFANALLEEVLTAKTVVTAPEAIRGTRV 147
　　　　　　 ||:: :|:|::| :::|　| ::|| :　:::::::|:::::|::::|
Db　　638 TLTLITV---VGYNKETVIKTINTNGENPYQPVSN-------LTATT-----QGQKV 679

Qy　　148 QGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV-EIKANGKRADFTETFESSTHGEA 206
　　　　　　:||:　|::|　| ::|:: 　::: 　:| :　:||: |||| :||:
Db　　680 TLKWDAPSTKTNATTN-TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWNDGS- 737

Qy　　207 PAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWN-----GMALNPDNYLISK 261
　　　　　　 :||::::|　:|| ::| :::|:|:|:|:|:|:|:|　::::　||||||||||||||||
Db　　738 -GYQILLDADHDQYGQVIPSDHTLWPNCSVPANLFAPFEYTVPENADPSCSPTNMIM-- 794

Qy　　262 DVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDTVVFEETPNGINKGGARFGLSTE 321
　　　　　　 ::　|:　|:|　::::::::::| 　|:|::| :||:　:::
Db　　795 DGTASVNI----------------PAGTY-------DFAI------------------ 811

Qy　　322 ADGAKPQS---VWI------ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP 371
　　　　　　 :: :　||　: ||
Db　　812 --AAPQANAKIWIAGQGPTKEDDYVFEAGKKY----HFLMKKMGSGDGTELTISEGGG- 863

Qy　　372 TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPV 431
　　　　　　 :|:: 　 :|　: 　:: :::::　|:
Db　　864 --SDYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGV 921

Qy　　432 QFNPVQN-LTGSAVGQKVTL 450
　　　　　　 |||||| :||::| ::|:
Db　　922 ADVTAQKPYTLTVVGKTITV 941

RESULT 6
AF2959
conserved hypothetical protein Atu3276 [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AF2959
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, G.; Grant, W.; Guenthner, C.; Kutyavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF2959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1052 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44092.1; PID:g17741659; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3276
A;Map position: linear chromosome

Query Match　　5.6%;　　Score 139.5; DB 2; Length 1052;
Best Local Similarity　20.9%;　　Pred. No. 0.12;
Matches　123;　Conservative　67;　Mismatches 159;　Indels 239;　Gaps 33;

Qy　　 13 TTTLSESFENGIPASWKTIDADGDGNNWTT------TPPPGGTSFAGHNSAICASSAS 64
　　　　　　 ::: ::|::| :||::::::| :　　　|||:|:|:::|:|
Db　　 52 TVTISGEAIIGRHSASEATV--TGDGSKWTTGDLQVGGDTSDPGG--LAGNGT------ 99

```
Qy  167 FRHFGCTDFFWINL-----DDVEIKANG---KRADFTETFESSTHGEAPAEWTTIDADGD 218
Db  476 SR-----SD--LNVGLYGNGSMTVEAGGAVKSRDGYVATYGGST-----SAVTVTGD 520

Qy  219 GQGWLCLSSQGLGWLTAHGGT-NVVASFSWNGMALNPDNYLI----SKDVTGA-TK 268
Db  521 GSSWAMTGTFFVGY-ASGATGNVTVS--NGGAIRATGVTLGDLAGASGTMTITGAGSK 575

Qy  269 VKYYY-------AVNDGFPG------DHY------AVMISKTGT 293
Db  576 VTAYVDNGTVNSGSVDVGFQGSGSLSVVNGGSLDAYNLYVGNALGSSGAVLVSVGCSHVS 635

Qy  294 -------NAGD------FTVVFEETPNGINKGGARFGLSTEADGAKPQSV 330
Db  636 VDGLMVVGNAGNGSVEITGGASLAAPTILIATEAGSTGVLSIGAGSGQTARSAGA------ 690

Qy  331 WIE-RTVDLPAGTKYVAFRHYNCSDLNYILLDDIQ------FTMGGSPTPTDYTY 378
Db  691 -VEARAIAFGAGNGSIVFNH--SETGYTLSADISGAGRVVAEAGVTTLSGNNS------ 740

Qy  379 TVYRDGTKIKEGLTETT-------FEEDG------VA 402
Db  741 --YSGGTTISAGMLKGTAKSFGSGGIVNNAELVVDGGGTLSNAISGTGSFEKTGDGNLLL 798

Qy  403 TGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLTGSAVGQKVTL 450
Db  799 TGNSTYSGATAVSAG-------KLSVNGSLASAVSVGSGATVGGTGTI 839
```

RESULT 8

B75622
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: B75622
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75622
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-691 <WHI>
A;Cross-references: GB:AE001826; PIDN:AAF12628.1; PID:g6460827; NID:g6460924; TIGR:DRB0
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRB0037
A;Map position: megaplasmid
A;Genome: plasmid
A;Note: plasmid MP1

```
Query Match             5.3%;  Score 132.5;  DB 2;  Length 691;
Best Local Similarity  21.9%;  Pred. No. 0.21;
Matches  111;  Conservative  46;  Mismatches  178;  Indels  171;  Gaps  23;

Qy   4 NPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTP----------PPGGTSF 51
Db  230 SPNPLPPTGTTT----------PGTGTGPVGSGNGPAESTGPAYTGIGSTGQPVVVDA 280

Qy  52 AGHNSAICASSASYINFEGP-------QNPD----NYLVTPELS-LPNGGTLTFWVCAQDA 100
Db  281 SGNQYAYPKADAETVNPDSVTMTTIITNPNTAATYELVPDLSGLPAGVTVTF----TDA 336

Qy 101 N-----YASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQ 153
Db  337 NGNPLPDTDGNGRPEVTAGPNGTA-----------TYRVVTYPDTESAAAVAG----- 380

Qy 154 KTVQLPAGTK----------YVAFRHFGCTDFFWINLDDVEIKANGKRADFT--- 195
Db  381 -PIRIPVGDGNRDGIVDATVTVNVLLSNLKFGNTNGTALGVSDVPVTRTVTQPSVTTAV 439
```

```
Qy  65 YINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGN------- 116
Db 100 -------LNVTAGGSV------DSTVA--HLGVVAGATGSAIVDGKGS 132

Qy 117 ---DASNF-----ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVA 166
Db 133 VWTVDRNSLEVGVSGAGSL--AVTGGGLVDAANIIIGTNTGG--NGSVRVSGADSTVK 186

Qy 167 FRHFGCTDFFWINL-----DDVEIKANG---KRADFTETFESSTHGEAPAEWTTIDADGD 218
Db 187 SR-----SD---LNVGLYGNGSMTVEAGGAVKSRDGYVATYGGST-------SAVTVTGD 231

Qy 219 GQGWLCLSSQGLGWLTAHGGT-NVVASFSWNGMALNPDNYLI---SKDVTGA-TK 268
Db 232 GSSWAMTGTFFVGY-ASGATGNVTVS--NGGAIRATGVTLGDLAGASGTMTITGAGSK 286

Qy 269 VKYYY------DHY------AVMISKTGT 293
Db 287 VTAYVDNGTVNSGSVDVGFQGSGSLSVVNGGSLDAYNLYVGNALGSSGAVLVSVGCSHVS 346

Qy 294 -------NAGD------FTVVFEETPNGINKGGARFGLSTEADGAKPQSV 330
Db 347 VDGLMVVGNAGNGSVEITGGASLAAPTILIATEAGSTGVLSIGAGSGQTARSAGA 401

Qy 331 WIE-RTVDLPAGTKYVAFRHYNCSDLNYILLDDIQ------FTMGGSPTPTDYTY 378
Db 402 -VEARAIAFGAGNGSIVFNH--SETGYTLSADISGAGRVVAEAGVTTLSGNNS 451

Qy 379 TVYRDGTKIKEGLTETT-------FEEDG------VA 402
Db 452 --YSGGTTISAGMLKGTAKSFGSGGIVNNAELVVDGGGTLSNAISGTGSFEKTGDGNLLL 509

Qy 403 TGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLTGSAVGQKVTL 450
Db 510 TGNSTYSGATAVSAG-------KLSVNGSLASAVSVGSGATVGGTGTI 550
```

RESULT 7

H98323
hypothetical protein AGR_L_3085 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: H98323
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markell, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: H98323
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1341 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK90114.1; PID:g15160106; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_3085
A;Map position: linear chromosome

```
Query Match             5.6%;  Score 139.5;  DB 2;  Length 1341;
Best Local Similarity  20.9%;  Pred. No. 0.16;
Matches  123;  Conservative  67;  Mismatches  159;  Indels  239;  Gaps  33;

Qy  13 TTTLSESFENGIPASWKTIDADGDGNNWTT-------TPPPGGTSFPAGHNSAICASSAS 64
Db 341 TVTSGEAIIGRHSASEATV--TGDGSKWTTGDLQVGGDTSDPGG--LAGNGT-------- 388

Qy  65 YINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGN------- 116
Db 389 -------LNVTAGGSV------DSTVA--HLGVVAGATGSAIVDGKGS 421

Qy 117 ----DASNF-----ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVA 166
Db 422 VWTVDRNSLEVGVSGAGSL--AVTGGGLVDAANIIIGTNTGG--NGSVRVSGADSTVK 475
```

```
Qy    196  ---ETPESSTH-----------GEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGT  239
                :         :    ||| ||||||||||||||||||||||||||||
Db    440  VFPMDLFNDGAYDGNYALSGSTPIGPVKYYATNPDTDGDG----VLSPAELAALPAE--  492

Qy    240  NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  299
                ||                    |        ||    || |||||||||||
Db    493  -IAS---------------TGAVPVKIEKTV-------YAVVTIPAGQAPGDYM        523

Qy    300  VVFEETPNGINKGGARFGLSTEADGAKPQ-SVWIERTVDLPAGT-----------      342
                            :  :  :  :    :  :: ||||||
Db    524  VT--QTATGSLSGTTKSFNTDKVTVTSPNGSLLIAKRVTTPGTTTPSLNATANPGDAVSY  581

Qy    343  KVYAFRHYNCSDLNYILLDDIQFTMG----------GSPTPTDYTYTV-----YRD-    383
                ||      ::     |     ::  :     :  :      ::
Db    582  TVTATNNYNTSLYGLVLRDPSSSNNLGSFSSSNVFGFIKPSSLRAITVSGVSGATVLYRTSN  641

Qy    384  -GTKIKEGLTE--TTFEEDGVATGNH      406
                |  |     |    ||| ||
Db    642  LNTWAAQPTVDANTTWVEVGVDTNNN  667


RESULT 9
A83412
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 31-Dec-2000
C;Accession: A83412
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83412
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-2468 <STO>
A;Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1874

Query Match             5.3%;  Score 131.5;  DB 2;  Length 2468;
Best Local Similarity   22.2%;  Pred. No. 1.3;
Matches  120;  Conservative  47;  Mismatches  184;  Indels  189;  Gaps  30;

Qy      5  PNPNPNPGTT----------TLSESFENGIPASWKTIDADGDGN----------   38
                |||||||       | |:  ||||
Db    754  PVINPSNGTTLSGTAEPGSSVTLTDG--NGNPIG--QVTADGSGNWSFTPSTPLADGTVV  809

Qy     39  NWTTTPPPGGTSFAGHN-----------SAICASSASYINFEGGPQNP-----   74
                  ||||                  :: | ||| :||||||
Db    810  NATATDPAGNTSGQGSTTVDGVAPTTPTVNLSNGSSLSGTAEPGSTVILTDGNGNPIAEV  869

Qy     75  ----DNYLVTPELSLPNGGTLTFWCAQDANYASEHYAVYASSTGNDASNFANALLEEV  129
                :|: |||| |  :|| ||||||||||||||||||||||||||||
Db    870  TADGSGNWTTYTPSTPIANGTVVN--VVAQDA---------AGN-SSPGASVTVDSQ    913

Qy    130  LTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANG  189
                :|||||                |   :       |  ||
Db    914  APAAPVNPS---NGTTLSGT--------AEPGAT----------VTLTD----GNG   945

Qy    190  KRADFTETFESSTHGEAPAEWTTIDADGDGQGW-----LCLSSGQLGWLTAH--GGTNV  241
                   |  :                        :  :  |||||||||||||
Db    946  NPIG---------------------QVTADGSG-NWSFTPGTPLANGTVVNATASDPTGNTSA  986

Qy    242  VASFSWNGMA--LNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTG----  292
                ||||||||    ||||||||||||||||||||||||||||||||||||||
Db    987  PASTTVDSVAPAAPVNPSN---GAEISGTAEPGATVTLTDGSGNPIGQVTADGSGNWSF  1043

Qy    293  ---TNAGDFTVV--FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAF  347
                || ||          ||||||||||||||||||||||||||
Db   1044  TPSTPLADGTVVNATATDPAG-NTGGQG---STTVDAIAPAT---PTVNLSNGSSLSG-  1094

Qy    348  RHYNCSDLNYILLDD------IQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGV  401
                :::|              :  ||     :     ::
Db   1095  ---TAEPGSTVILTDGNGNPIAEVTADGS--GNWTYI---PSTPIANGTVVNVVAQD--  1143

Qy    402  ATGNHEYCVEVKYTAGVSPKECVN----------VTVDPVQFNPVQNLTGSAVG       445
                ||                |  :  ||||||||||
Db   1144  ASGNSSPPATVTVDSSAPPAPVINPSNGVVISGTAEAGATVTLTDAGGNPIGQVTADGSG  1203


RESULT 10
C89874
autolysin [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: C89874
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, N.; Inoue, R.; Kaito, C.; Sekimizu, K.;
  C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: C89874
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1248 <KUR>
A;Cross-references: GB:BA000018; PID:g13700854; PIDN:BAB42150.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: atl

Query Match             5.2%;  Score 128.5;  DB 2;  Length 1248;
Best Local Similarity   22.3%;  Pred. No. 0.84;
Matches  105;  Conservative  47;  Mismatches  165;  Indels  153;  Gaps  26;

Qy      3  PNPNPNPNPGTTT--LSESFENGIPASWKTIDADGDG--NNWTTTPPP---------   46
                | ||||:||     :|: :      :      :  :
Db    586  PTPTPIFKPSTPTPTNNKJTVSSLNGV--AQINAKNNGLFTTVYDKTGKPTKEVQKTFA  641

Qy     47  ------GGTSF--AGHNSA-----ICASSASYINFEGPQN-PDNVLVTPE--LSLPN   87
                  |||           :   : :  :  ::     :
Db    642  VTKEASLGGNKPFLVKDYNSPTLIGWVKQGDVIYNNAKSPVNVMQTYVKPGTKLYSVPW  701

Qy     88  G------------GTLTFWVCAQDANYASEHYAVYASSTGNDASNFAN---------ALLE  127
                               :     :      :  :         ::
Db    702  GTYKQEAGAVSGTGNQTFKATKQQQIDKS---IYLFGTVNGKSGWVSKAYLAVPAAPKK  757

Qy    128  EVLTAKT----VVTAPEAIR----GTRVQGTWYQKTVQLPAGTKYVAFRHFG        171
                :|                   |   :       |   |
Db    758  AVAQPKTAVKAYTVKPQTTQTVSKIAQVKPNNTGIRASVYEKTAK--NGAKY-ADRTFY  814

Qy    172  CT--------------DFFWINLDDVEIKANGKRADFTE--TFESSTHGEAPAE       209
                               :  :                     :  ::
Db    815  VTKERAHGNETYVLLNNTSHNIPLGWFNVKDLNVQNLGKEVKTTQKYTVNKSNNGLSMVP  874

Qy    210  WTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWN------GMALNPDNYLISKDV  263
                |                         :           :
Db    875  WGT----KNQVILTGNNI--AQGTFNATKQVSVGKDVYLYGTINNRTGWVNAKDL      923

Qy    264  TGATKVK--------YYYAVNDGFPGDHYAVMISKTGT-------NAGDFTVVFEETPN  307
                |             ||||||||||||||||||||        ||
Db    924  TAPTAVKPTTSAAKDYNYTVIKNG-NGYYYVTPNSDTAKYSLKAFNEQPFAVVKEQVIN  982

Qy    308  G------INKGGARFGLSTEADGAK------PQSVWIERTVDLPAGTKY           344
                       | |||| |                 :  ::
Db    983  GQTWYYGKLSNGKLAWIKST--DLAKELIKYNQTGMTLNQVAQIQAGLQY         1030


RESULT 11
AH2515
hypothetical protein alr7304 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120al
C;Species: Nostoc sp. PCC 7120
```

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2515
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-4936 <KUR>
A;Cross-references: GB:BA000020; PIDN:BAB78388.1; PID:g17135842; GSPDB:GN00180
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr7304
A;Genome: plasmid

Query Match   5.2%;   Score 128; DB 2; Length 4936;
Best Local Similarity   20.4%;   Pred. No. 5.5; Indels 190; Gaps 27;
Matches 108; Conservative 60; Mismatches 171; Indels 190; Gaps 27;

```
Qy      1 GTPNPNPNPNPGT------------TTLSESFENGIPASWKTIDADGDGNNWTT----TPPPGG     48
              :  ||||  ||:  | ||||||  |||||||
Db   3161 GLSNFN---TPGTYQLQVDATTVQDNAGN----------------SGDAARTTTFTIAAPPTGV   3206

Qy     49 TSFAGHNSAICASSASYINFEGPQNPDNYLV-------TPELSLP-NGG----------TLTF     93
            ||||||      |::   ::||||     :   |||  |  |||
Db   3207 T----ITQSGGSTAVIEG-GNTDSYTLVLRTQPTADVTVTVLNTGSQITTDKTTLTF   3257

Qy     94 --WVCAQ---------DANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEA    141
               ||   |         :    ::|||||||||||||||:
Db   3258 TSANWNTPQTITVNAVNDTITEGNHTSTISHSISSTDTNYSNVTLPDI--AVSITDNDAE   3315

Qy    142 IRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVBIKANGKRA--DFTETFE    199
            |  :||||||||||||||||||||||||||||||||||||||||||  ::|
Db   3316 IRGMK---------------WNDIDGDGVKDTGEPGLQGWTIYLD   3345

Qy    200 SSTHGEAP--AEWTTIDADGDGQWLCLSSG------------QLGWLTAHGGTNV--------    241
             |:|:      |||||||||||||:||| |
Db   3346 SNTNGQLDNGEISTTTDANGNYQ FTNLRPGVYTVAEVQQPGWKQTPPGTNITTTNADIP    3404

Qy    242 -----VASFSWNGMALN---PDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKT    291
                 :    |:||  || ||||||||||||||||||||||||  ::|
Db   3405 LAIPSLDMISPGDSNGIQLNFSAANYIVKEDGTAITEV---WVTRTGNTSSAVSATLSFT   3461

Qy    292 -GTNAG-----------------------DFTVVFEE-----------------TPNGIN----    310
                 |||                            | |
Db   3462 DGTATGCGCGASSVNNDFNNVPFTIAFAENETSKLISVQNALLANPNAIKIRNDSKVEGN   3521

Qy    311 -------KGGARFGLSTEA------DGAKPQSVWIERTVDLPAG---TKYVAFRHYNCS    353
                   |||||||||          ||:|||||||   | :  :|||||||
Db   3522 EYFTIKLTNPTGGAVIGNQSIATVTIIDDEAPSDITVTPPLETPSTTITSAVDSQAIYLI   3581

Qy    354 DINYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVA    402
             |||:||||||||||||||||||||||||||||||||||||  :|
Db   3582 NLNNFWADSRFANIKGN---DFTSVIIDTGIDLNHPFGADTDNNGIA   3626
```

RESULT 12
S10789
amylase A-180 – alkaliphilic eubacterium 163-26
C;Species: alkaliphilic eubacterium 163-26
C;Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 15-Oct-1999
C;Accession: S10789
R;Candussio, A.; Schmid, G.; Boeck, A.
Eur. J. Biochem. 191, 177-185, 1990
A;Title: Biochemical and genetic analysis of a maltopentaose-producing amylase from an a
A;Reference number: S10789; MUID:90336627; PMID:1696201
A;Accession: S10789
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1684 <CAN>

A;Cross-references: EMBL:X53373; NID:g48305; PIDN:CAA37453.1; PID:g48306
A;Accession: AH2515

Query Match   5.1%;   Score 127.5; DB 2; Length 1684;
Best Local Similarity   20.9%;   Pred. No. 1.5; Gaps 27;
Matches 102; Conservative 53; Mismatches 184; Indels 149; Gaps 27;

```
Qy     23 GIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA-SSASYINFEGPQNPDNYLVTP     81
            ||||||||||||||||||||         :||:: |::
Db    204 GLPRDWTPNQA--QQQNWHT----------------HNDIMNKDNEAAWANWWGSD----    241

Qy     82 ELSLPNGGTLTFWVCAQD-ANYASEHYAVYASSTGNDASNFANALLEEVLTAKTV-VTAP    139
                                    ||:       :::
Db    242 --------------WIRADETAGY----------DNCGGSEQTMCIGFLPDIKTEVTTGVDLP    280

Qy    140 EAIRGT-RVQGTWYQKTVQLPAGTKYVAFRHFGCTDFF--WINLDDVEIKANGKRADFTE    196
             |||||          |                          ::|:
Db    281 PILRNKWNDQASGYEDWF-VPAAEPYRQDLNIAPKDYLIKWITSWEEFGIDGPRVDTAK    339

Qy    197 TFESSTHGEAPAE----------WTTIDADGDGQWLCLSSGQLGWLTA-------HG----    237
               :|                    ::||||
Db    340 HVEIERWAELKNEARVALQTWRENNPDKPGANW---DDNFWMTAEVFGHGLGKSEYFD    394

Qy    238 -GTNVVVASFSW---------------NGMALNPDNYLISKDVTGATKVKYYYAVNDGFP    280
                     |||:               ||:  ::||
Db    395 FGFDSVINFBFQNANFNNLBGLFSRYANSINTDPFDFNMLSYVSSHDTKL--YSRDD---    448

Qy    281 GDHYAVMISKTGTN---AGDFTVVF-EETPNGINKGGARFGLSTEADGAKPQSVW--IE    333
               |||||          |||:|||
Db    449 --------LIQAGTALLLLPGGVQVFYGDETARPLGDGG----SDPEQTRSSMNWANIN    496

Qy    334 RTVDLPAGTKYVAPR-----------LLDDIQFTMGGSPTPT    374
                |  | |:                      :::
Db    497 QNV-LSHWQKLGQFRNNHIAIGAGAHQKLSDSPYTFARTYESDDIVDEVVVATGAQGTTA    555

Qy    375 DYTYTVYRDGTKIKEGLT--ETFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPVQ    432
               :|||||||||||||
Db    556 VTVEGVFEDGTVVRDAYTGDETTVTK-GTAT-------FTAGTQGIILIENTAEPVT    604

Qy    433 FNPVQNLT    440
             |:::|
Db    605 NLPIVSAT    612
```

RESULT 13
B75489
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: B75489
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-713 <WHI>
A;Cross-references: GB:AE001925; GB:AE000513; NID:g6458383; PIDN:AAF10267.1; PID:g64583
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0685
A;Map position: 1

Query Match   5.1%;   Score 127; DB 2; Length 713;
Best Local Similarity   21.6%;   Pred. No. 0.51;
Matches 129; Conservative 47; Mismatches 214; Indels 208; Gaps 29;

```
Qy      2 TPNPNP-NPNPGTTLSE------SFENGIPASWKTIDADGDGNNWTTTPP---PGGT----     49
            |||| | ||||||               |||||||||||||||||||||   ||||
Db     86 TPESTPTNPNPPSETIETPPVITVVNPVPSFTIT---PNDGSKDTTQPDYTNPGQTREVK    142
```

```
Qy    50  -----SPAGHNSAICASSASYINFE-----GPQNPDNVLVTPE----LSLPNGGTL--  91
                 ::  |||   :||            |:       ||::  |||
Db   143  PCDKNVTF-----AYTLTNTGNVNGESYTLTNTPDPTGAVKTPENIRFYLDSNNNGQLDQ 197

Qy    92  -----TFWVCAQDANYASEHYAVY------ASSTGNDASNFAN--ALLE 127
                ||:|         |:||:|          |::||   :|:|
Db   198  SEIAAGATTTITDVAINQTVKFQVYDVPCTATSTDKFGGDPTGTRNDNPNFSNDPTLPR 257

Qy   128  EVLTAKTV------TAPEAIRGTRVQGTWYQKTV-QLPA 160
               :                   |:|  :||||
Db   258  DANNSNTVTINRKDGVVIGPKADPDGNGNPVTPAYNSPEGINIVPTASDTQVATVTTLPT 317

Qy   161  GTKYVAFRHFGCTDFFWINLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQ 220
          |||||||          :||:                |:   ||::   :||
Db   318  SGVTVTFN------------TIQNTGNR-----TDTFELTQINTFPAGTTVVFKDANGN 360

Qy   221  GWLCLSSGQLGWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFP 280
             ||:      :|:|      ||   |   ||        |:|
Db   361  A------------LPTVNGKPVVQNVPENGTV---DIQVTLLPAGVTPTQL-----SQP 401

Qy   281  GDHYAVMISKTGTN---AGDFT--VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERT 335
               :::        :|||      :|  |    |  |||
Db   402  ---AVTVTTTSQNDPVPKSDTTKDIIEVKVP-----GIAFGDPTPGLGGDPTPVGTPPT 451

Qy   336  -----VDLPAG---TKYVAFRHYNCSDLNYTILLDDIQFTMGG-----SPT 372
              ||                          |||
Db   452  GVPGNPGTPLTPGNPQTCTAPIRTYLPMEIANLGSQDDAFVVSGTAPVTVLNPDGTVNPT 511

Qy   373  PTDYTYTVYRD---GTKIKEGLTETFFEEDGVATGNHEYCVEVKY----- 414
            :|                            ||
Db   512  PVIVPVVYYRDVNGDGKLDAGDTEL-----QGGNTGTIKPGEEIKLIAVVDVPCAAAQTI 567

Qy   415  -----TAGVSPKE-----CVNVTVDPVQFNPVQNLTGSAVGQKVT 449
              :||                 |||   ::|::   ||::
Db   568  TLNQEAKSPTTGVSQKDPNDTITVGGNGGKPIVTKTVDKATANPGDTLTYTIIGKNTS 625
```

RESULT 14
A86685
prophage pil protein 46, tail component [imported] - Lactococcus lactis subsp. lactis (s
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: A86685
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: A86685
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1441 <STO>
A;Cross-references: GB:AE005176; PID:g12723361; PIDN:AAK04579.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: pil46

```
Query Match       5.1%;   Score 126;   DB 2;   Length 1441;
Best Local Similarity  22.3%;   Pred. No. 1.5;
Matches  98;  Conservative  49;  Mismatches  133;  Indels  160;  Gaps  26;

Qy    30  TIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYINFEGPQNP----DNYLVTPELSL  85
                                |::|             ||       |||
Db   602  TISSSG---------TVTPTTGWTS-----------QVPTLVKEQYL----------- 628

Qy    86  PNGGTLTFWVCAQDANYASEHYAV-YASSTGN-----DASNFANALLEEVLTAKTV 135
                     ||:|         |::|
Db   629  ---WTKTLWTYTD--NTSETGYSVSYIAKDGNNGHDGFPGKDGVGLISNTIIEYV----- 677

Qy   136  VTAPEAIRGT-RVQGTWYQKTVQLPAGTKYVAFRHF-----GCTDFFWINL----- 180
                        |:|  :||||        ::|
Db   678  -GAVSGTSKPTGGWSTTIPTVPAG-QYLWTRTTWQYTDGTSEQGYINALMGLITGASG 732

Qy   181  -DDVEIKANGKRADFTE-TFESSTHG-EAPAE----------WTTIDADG 217
```

```
Db   733  RDGIAGK-DGKGIKATATITYQASTNGTTAPTGTWSTSVPSVAKGSFLWTRTIWTWTD--- 788
                  ||                                        :||
Db   218  DGQGWLCLSSGQLGWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVN- 276
            :  |                        ||
Db   789  -----NTTETGYAVAYMGTN-------------GNNGTNGIAGKDGTGIKTTTITYAVGT 830

Qy   277  --------------DGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEA 322
                         ::  :|        |||       ::|||           ||
Db   831  SGTTAPTGGWNSQVPNVPAGQY--LWTKT--VWDYTDKTSETGYSVSKFGEK-GDKGDQ 884

Qy   323  DGAKPQSVWIERTVDLPAG------TKYVAFRHYNCSD--LNYILLDD-----IQFTMG 368
              ||                        ||
Db   885  GVQGIQGVDGRQGIPGPKGADGKTQYTHIAYANSADGKTNFSTSDSNRTYIGMYVDFNIN 944

Qy   369  GSPTPTDYTYTVYR--DGTK 386
             |||::|:||:    |:|
Db   945  DSTTPSDYSWTLVKGADGTQ 964
```

RESULT 15
C86622
hypothetical protein yqbK [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: C86622
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: C86622
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1649 <STO>
A;Cross-references: GB:AE005176; PID:g12724583; PIDN:AAK05677.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yqbK

```
Query Match       5.1%;   Score 126;   DB 2;   Length 1649;
Best Local Similarity  20.4%;   Pred. No. 1.8;
Matches  110;  Conservative  56;  Mismatches  189;  Indels  184;  Gaps  28;

Qy    46  PGGTSFAGHNSAICASSASYINFEG-PQN----------PDNY----------LVTPELSLPN  87
            ::                :|        |||
Db    65  PSGTTFNG------------NSVSQPNRTPLPQNSVEDAIMPNSFGWQQTIANLTLPTNQVTAVQ 117

Qy    88  GGTLTFWVCAQDANYASEHYAVYASSTGNDASNF-------ANALLEEVLTAKTVVTAPEA- 141
                                        ::|||   ::|
Db   118  NGTIIGSYPTTDAGVQQALYAMY-SATNASTSDFGLYYGANTTLSNSL-LKAVVSNPSAT 175

Qy   142  -----------IRGTRVQGTWYQKTVQL-----PAGTKYVAFR----HFGCTDFFWINLDDV 183
                :                               ||
Db   176  NMTFSTLKGHAKSLTWISNPADLLTSSNSSQPSGTNYTGTLPSNVYFGVPTIF---RNV 231

Qy   184  EIKANGKRADFTETFESSTHGEAPAEWTT-----------IDADGDGQGW 222
              :||                                :|
Db   232  TVAASGDNIYAQGNAFATNG-----SWITGAPNIYGGTDNSDISGNTNLYIGATGSIAGW 287

Qy   223  LCLSSGQLGWLTAHGGTNV--------VASFSWNGMALNPDNYLLISKDVTG--ATKV 269
            |                          ||
Db   288  -NIYGGNASAATISGNTHVTIAQSSTINSVTGGSASGTTISGNTNL--DISGAIASQI 343

Qy   270  KYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNG-----------INKGGARFG--LS 319
                          ||                ||
Db   344  TNIYGAGIG---------TSNSPVNVNGNVTTYVNSTNGGARYQLYQGGTVVGNISSGSIY 394

Qy   320  TEADGAKPQSVWIERTVDL-----------PA-----------GTKYVAFRHYNCSDLNYIL 359
                                                        ::
Db   395  NTLSGA---GGWTGATSNINGAGGPASTFNGGSFQGNIGTSGAGNVISNSYNTSSFT--- 448

Qy   360  LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTIKEGLTETFPEEDGVATGNHEYCVEVKYTAG--- 417
```

```
Db   449 TGQALFTGGNAGTGASYAQA----------TNSTTAAQGILYANITNYIKSAFTTGTA 496
                            :||:
Qy   418 --------------VSPKE---------CVNVTVDPVQFNPVQNLTGSAVGQKVT 449
             ||||| |   : :| : ||:|
Db   497 GAVYGIVGGNGHDSLKISPSQWGLGSQTGLDSAVGVT-DAKAYGQIPSTTVVSNAQKIT 554
```

Search completed: May 18, 2004, 11:47:49
Job time : 13.906 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:35:14 ; Search time 8.3342 Seconds
                 (without alignments)
                 2848.981 Million cell updates/sec

Title: US-08-570-311-18
Perfect score: 2480
Sequence: 1 GTPNPNPNPGTTTLSESF.........QNLTGSAVGQKVTLKWDAPN 456

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2480	100.0	2628	1	HGA2_PORGI	Q51845 porphyromon
2	2441	98.4	2164	1	HGA1_PORGI	P59915 porphyromon
3	1453	58.6	989	1	PRTH_PORGI	P46071 porphyromon
4	266	10.7	991	1	CPG1_PORGI	P28784 porphyromon
5	131.5	5.3	1256	1	ATL_STAAU	P52081 staphylococ
6	126	5.1	721	1	OGP_MOUSE	Q62010 mus musculu
7	121.5	4.9	465	1	SLAP_LACBR	Q05044 lactobacill
8	121	4.9	1090	1	GUXB_CELFI	P50899 cellulomona
9	118.5	4.8	1385	1	C5AA_BACUD	Q45760 bacillus th
10	114	4.6	872	1	GUXA_CELFI	P50401 cellulomona
11	113	4.6	2660	1	YEEJ_ECO57	Q8X8V7 escherichia
12	112.5	4.5	1034	1	BGAL_BACME	O52847 bacillus me
13	112.5	4.5	1260	1	ALS1_CANAL	P46590 candida alb
14	110.5	4.5	1122	1	ADP1_MYCGA	Q49379 mycoplasma
15	109.5	4.4	331	1	PME_ASPAC	Q12535 aspergillus
16	109.5	4.4	607	1	YK28_ARCFU	O28251 archaeoglob
17	108.5	4.4	1953	1	BIGA_SALTY	P25927 salmonella
18	108	4.4	699	1	CHI1_BACCI	P20533 bacillus ci
19	108	4.4	987	1	YD94_METJA	Q58789 methanococc
20	108	4.4	1289	1	C5AB_BACUD	Q45753 bacillus th
21	107.5	4.3	551	1	AMYB_THETU	P19584 thermoanaer
22	107.5	4.3	1200	1	HYAL_STRPU	O76536 strongyloce
23	107	4.3	1656	1	OMPB_RICJA	O06653 r outer mem
24	107	4.3	2249	1	OMPA_RICRI	P15921 rickettsia
25	106.5	4.3	1122	1	ADP2_MYCGA	Q9REM8 mycoplasma
26	106.5	4.3	2124	1	PGCA_RAT	P07897 rattus norv
27	106.5	4.3	2358	1	YEEJ_ECOLI	P76347 escherichia
28	106	4.3	1045	1	PRTS_SERMA	P09489 serratia ma
29	106	4.3	1300	1	120K_RICRI	P14914 rickettsia
30	106	4.3	1654	1	OMPB_RICRI	Q53047 r outer mem
31	105.5	4.3	1220	1	C5AC_BACTU	P56995 bacillus th
32	105.5	4.3	1534	1	YFAS_ECO57	Q8XE35 escherichia
33	105	4.2	827	1	XANP_XANS2	Q60106 xanthomonas
34	105	4.2	1045	1	GUNB_CELFI	P26225 cellulomona
35	105	4.2	2812	1	ZAN_HUMAN	Q9Y493 homo sapien
36	103.5	4.2	837	1	XYNZ_CLOTM	P10478 clostridium
37	103.5	4.2	1655	1	OMPB_RICCN	Q9KA3 r outer mem
38	103.5	4.2	1861	1	APU_THETU	P38536 t amylopull
39	103	4.2	1157	1	C8AA_BACUK	Q45704 bacillus th
40	102.5	4.1	282	1	PRTA_ASPNG	P24665 aspergillus
41	102.5	4.1	1045	1	PRTT_SERMA	P29805 serratia ma
42	102	4.1	435	1	AM3D_ORYSA	P27933 oryza sativ
43	102	4.1	941	1	GUN_BACS6	P19424 bacillus sp
44	101.5	4.1	1462	1	PTP6_DROME	P16620 drosophila
45	101.5	4.1	2132	1	PGCA_MOUSE	Q61282 mus musculu

ALIGNMENTS

RESULT 1
HGA2_PORGI
ID   HGA2_PORGI   STANDARD;   PRT;   2628 AA.
AC   Q51845;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97047672; PubMed=8926061;
RA   Han N., Whitlock J., Progulske-Fox A.;
RT   "The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381
RT   contains four large, contiguous, direct repeats.";
RL   Infect. Immun. 64:4000-4007(1996).
CC   -!- FUNCTION: Agglutinates erythrocytes.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U41807; AAB17128.1; -.
DR   PIR; T28651; T28651.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT   SIGNAL   1   24   POTENTIAL.
FT   CHAIN   25   2628   HEMAGGLUTININ A.
FT   DOMAIN   25   539   PEPTIDASE C25-LIKE 1.
FT   DOMAIN   540   995   PEPTIDASE C25-LIKE 2.
FT   DOMAIN   996   1451   PEPTIDASE C25-LIKE 3.
FT   DOMAIN   1452   1907   PEPTIDASE C25-LIKE 4.
FT   DOMAIN   2074   2628   PEPTIDASE C25-LIKE 5.
SQ   SEQUENCE   2628 AA;   283324 MW;   61C4DE32540C99DA CRC64;

Query Match   100.0%;   Score 2480;   DB 1;   Length 2628;
Best Local Similarity   100.0%;   Pred. No. 1.6e-163;
Matches   456; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA 60

Db   950 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA 1009

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Db   1010 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1069

```
Best Local Similarity   98.9%;   Pred No. 6.2e-161;
Matches  449;   Conservative   2;   Mismatches   3;   Indels   0;   Gaps   0;

Qy      3 PNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS  62
          |||||||||||||||||||||||||||||||||||||||::|||||||||||||||||
Db    496 PNGTPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGSSFAGHNSAICVSS 555

Qy     63 ASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFA 122
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    556 ASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFA 615

Qy    123 NALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDD 182
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    616 NALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDD 675

Qy    183 VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVV 242
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    676 VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVV 735

Qy    243 ASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVF 302
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    736 ASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVF 795

Qy    303 EETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDD 362
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    796 EETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDD 855

Qy    363 IQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKE 422
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    856 IQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKE 915

Qy    423 CVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||
Db    916 CVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 949
```

```
RESULT 3
PRTH_PORGI
ID   PRTH_PORGI   STANDARD;         PRT;       989 AA.
AC   P46071;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protease prtH (EC 3.4.22.-).
GN   PRTH.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=95012612; PubMed=7927685;
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RT   "Cloning and characterization of a new protease gene (prtH) from
RT   Porphyromonas gingivalis.";
RL   Infect. Immun. 62:4279-4286(1994).
RN   [2]
RP   ERRATUM.
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RL   Infect. Immun. 62:5707-5707(1994).
CC   -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
CC       P.GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
CC       IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
CC       AND IS A VIRULENCE FACTOR.
CC   -!- SUBCELLULAR LOCATION: In membrane vesicles.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
Qy    121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 1129

Qy    181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 1189

Qy    241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 1249

Qy    301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1309

Qy    361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1369

Qy    421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db   1370 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 1405
```

```
RESULT 2
HGA1_PORGI
ID   HGA1_PORGI   STANDARD;         PRT;      2164 AA.
AC   P59915;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA OR PG1837.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=22829867; PubMed=12949112;
RA   Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA   Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
RA   Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
RA   Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA   Dewhirst F.E., Fraser C.M.;
RT   "Complete genome sequence of the oral pathogenic bacterium
RT   Porphyromonas gingivalis strain W83.";
RL   J. Bacteriol. 185:5591-5601(2003).
CC   -!- FUNCTION: Agglutinates erythrocytes (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC       Complete proteome.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   2164       HEMAGGLUTININ A.
FT   DOMAIN       26    539       PEPTIDASE C25-LIKE 1.
FT   DOMAIN      540    991       PEPTIDASE C25-LIKE 2.
FT   DOMAIN      992   1443       PEPTIDASE C25-LIKE 3.
SQ   SEQUENCE   2164 AA;  233387 MW;  6DFAB2283258EC63 CRC64;

DR   EMBL; AE017178; AAQ66831.1; ALT_INIT.
DR   TIGR; PG1837; -.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
```

Query Match   98.4%;   Score 2441;   DB 1;   Length 2164;

CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L27483; AAA51298.1; -.
DR   HSSP; P23882; 1FMT.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR002376; formyl_transf.
DR   Pfam; PF00551; formyl_transf; 1.
KW   Hydrolase; Thiol protease; Repeat; Virulence.
FT   REPEAT   270   323
FT   REPEAT   528   581
SQ   SEQUENCE   989 AA;   110238 MW; FA85FE8A3AC8944C CRC64;

Query Match
Best Local Similarity   58.6%; Score 1453; DB 1; Length 989;
Matches 276; Conservative 11; Mismatches 30; Indels   8; Gaps   3;

Qy   136 VTAPEAIRGTRVGGTWYQKTVQLPAGTKYVAFRHFG----CTDFFWINLDDVEIKANGKR 191
              :||||||| |||: ||||:|||||||| ||:::| .: || ||: | ||||
Db   34 VRSPEAIRG-RIQGTWRQKTVDLPAGTEICCFPSLPKAPICSTSTLMRLI---RSKTNAKR   89

Qy   192 ADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWNGMA 251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   90 ADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMA 149

Qy   252 LNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINK 311
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   150 LNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINK 209

Qy   312 GGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP 371
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210 GGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP 269

Qy   372 TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPV 431
              ||||||||||||||||||||||||||||||||||||||||:|||||||||: ||||||:
Db   270 TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTVNST 329

Qy   432 QFNPVQNLTGSAVGQKVTLKWDAPN 456
              |||||: :||:|||||: |||
Db   330 QFNPVKNLKAQPDGGDVVLKWEAPS 354

RESULT 4
CPG1_PORGI
ID   CPG1_PORGI   STANDARD;   PRT;   991 AA.
AC   P28784; Q45168;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
DE   (RGP-1).
GN   RGPA OR RGP1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 228-290 AND 517-541.
RC   STRAIN=381;
RX   MEDLINE=95168884; PubMed=7864651;
RA   Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
RA   Ikehara Y.;
RT   "Structural characterization of arginingipain, a novel
RT   arginine-specific cysteine proteinase as a major periodontal
RT   pathogenic factor from Porphyromonas gingivalis.";
RL   Arch. Biochem. Biophys. 316:917-925(1995).
RN   [2]
RP   SEQUENCE OF 228-270.
RC   STRAIN=HG66;
RX   MEDLINE=92406812; PubMed=1527017;
RA   Chen Z., Potempa J., Polanowski A., Wikstrom M., Travis J.;
RT   "Purification and characterization of a 50-kDa cysteine proteinase
RT   (gingipain) from Porphyromonas gingivalis.";

RL   J. Biol. Chem. 267:18896-18901(1992).
CC   -!- FUNCTION: Thiol protease which is believed to participate in
CC   intracellular degradation and turnover of proteins. Its
CC   proteolytic activity is a major factor in both periodontal tissue
CC   destruction and in bacterial host defense mechanisms. Activates
CC   complement C3 and C5.
CC   -!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
CC   immunoglobulins, with a preference for Arg in P1, and hydrophobic
CC   residues in P2 and P3.
CC   -!- ENZYME REGULATION: Requires cysteine for activation and Ca(2+)
CC   and/or Mg(2+) for stabilization. It is stimulated by glycine-
CC   containing dipeptides. It is resistant to inhibition by proteinase
CC   inhibitors in human plasma.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; D26470; BAA05484.1; -.
DR   PIR; I40229; I40229.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
KW   Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
FT   SIGNAL   1   24
FT   PROPEP   25   227
FT   CHAIN   228   991
FT   ACT_SITE   438   438   PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   471   471   NUCLEOPHILE (BY SIMILARITY).
FT   CONFLICT   264   265   RT -> TK (IN REF. 2).
SQ   SEQUENCE   991 AA; 108782 MW; 03EE3F43CBBE2544 CRC64;

Query Match
Best Local Similarity 10.7%; Score 266; DB 1; Length 991;
Matches 114; Conservative 50; Mismatches 164; Indels 112; Gaps 19;

Qy   30 TIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYINPEGPQNPDNYLVTPELSLPNGG   89
              ::: :   .:::   :|:   :.:
Db   595 SVNVSCDYNGAIATISANGKMF---GSAVVENGTATINLTG------------LTNES 637

Qy   90 TLTFWVCAQDANYASEHYAVYASSTG--NDASNFANALLEEVLTAKTVVTAPEAIRGTRV 147
              |||| :   .:   : .: :: .: |::.|
Db   638 TLTLTV----VGYNKETVIKTINTNGEPNYQPVSN------LTATT---------QGQKV 679

Qy   148 QGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV-EIKANGKRADFTETFESSTHGEA 206
              || ||||||| |||||||||| :: .| ||.   : |: :
Db   680 TLKWDAPSTKTNATTN-TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWNDGS- 737

Qy   207 PAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWN----GMALNPDNYLISK 261
              :.:   :::   :.|::   ::
Db   738 -GYQILLDADHDQYGQVIPSDTHTLWPNCSVPANLFAPFEYTVPENADPSCSSPTNMIM-- 794

Qy   262 DVTGATVKVKYYYAVNDGFPGDHYAVMISKTGNAGDTVVFEETPNGINKGGARFGLSTE 321
              : | .::.:
Db   795 DGTASVNI----------------------PAGTY---------------DFAI------ 811

Qy   322 ADGAKPQS---VWI------ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP 371
              :: :   :|: || : .   .:|:
Db   812 ---AAPQANAKIWIAGGFPTKEDDYVFEAGKKY----HFLMKKGMGSGDGTELTISEGG- 863

Qy   372 TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPV 431
              ||:|||||||||||||||||||||||:||||||||||||||||: ||||||:
Db   864 --SDYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGV 921

```
Qy       432  QFNPVQN-LTGSAVGQKVTL 450
              |:|:.|:
Db       922  ADVTAQKPYTLTVVGKTITV 941


RESULT 5
ATL_STAAU
ID   ATL_STAAU     STANDARD;      PRT;   1256 AA.
AC   P52081;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Bifunctional autolysin precursor [Includes: N-acetylmuramoyl-L-alanine
DE   amidase (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
DE   acetylglucosamidase (EC 3.2.1.96)].
GN   ATL.
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
RC   STRAIN=RN450;
RX   MEDLINE=95116542; PubMed=7816834;
RA   Oshida T., Sugai M., Komatsuzawa H., Hong Y.-M., Suginaka H.,
RA   Tomasz A.;
RT   "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
RT   alanine amidase domain and an endo-beta-N-acetylglucosaminidase
RT   domain: cloning, sequence analysis, and characterization.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 8325-4;
RA   Foster S.J.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOBIOSYL UNIT IN
CC       HIGH-MANNOSE GLYCOPEPTIDES AND GLYCOPROTEINS CONTAINING THE
CC       -[(MAN)5(GLCNAC)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
CC       RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
CC       OLIGOSACCHARIDE IS RELEASED INTACT.
CC   -!- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
CC       residues and L-amino acid residues in certain bacterial cell-wall
CC       glycopeptides.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC       unit in high-mannose glycopeptides and glycoproteins containing
CC       the -[Man(GlNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC       residue remains attached to the protein; the rest of the
CC       oligosaccharide is released intact.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
CC       EXTRACELLULAR LYTIC ENZYMES.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
CC       ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
CC       GLYCOSYL HYDROLASES.
```

```
DR   EMBL; D17366; BAA04185.1; -.
DR   EMBL; L41499; AAA99982.1; -.
DR   InterPro; IPR002502; Amidase_2.
DR   InterPro; IPR002901; Amidase_4.
DR   Pfam; PF01510; Amidase_2; 1.
DR   Pfam; PF01832; Amidase_4; 1.
DR   SMART; SM00644; Ami_2; 1.
DR   SMART; SM00047; LYZ2; 1.
KW   Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30   1256       BIFUNCTIONAL AUTOLYSIN.
FT   DOMAIN      199    775       N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
FT   DOMAIN      776   1256       ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
FT   REPEAT      425    589       1.
FT   REPEAT      596    758       2.
FT   REPEAT      770    932       3.
SQ   SEQUENCE   1256 AA;  137384 MW;  2BB76CAA292FDD20 CRC64;


Query Match             5.3%;  Score 131.5;  DB 1;  Length 1256;
Best Local Similarity  22.3%;  Pred. No. 0.24;
Matches  105;  Conservative  47;  Mismatches 165;  Indels 153;  Gaps  26;

Qy      3 PNPNPNPGTTT------LSESFENGIPASWKTIDADGDG---NNWTTTPPP-------- 46
           |||                ||||||                   ||:||
Db    594 PTPTPTPKPSTPTTNNKLTVSSLNGV----AQINAKNNGLFTTVYDKTGKPTKVQKTFA 649

Qy     47 ------GGTSF---AGHNSA----ICASSASYINFEGPQN-PDNYLVTPE---LSLPN 87
                  :||      ||||          |||  |:  :| |:  :|::.:  :.:|
Db    650 VTKEASLGGNKFYLVKDYNSPTLIGWKQGDVINNAKSPVNVMQTYTVKPGTKLYSVPW 709

Qy     88 G-------------------GLTFWVCAQDANYASEHYAVVASSTGNDASNFAN------ALLE 127
           :                    :| | | |    |:      || |:||    |  |: || | ||
Db    710 GTYKQEAGAVSGTGNQTFKATKQQQIDKS---IYLFGTVNGKSGWVSKAYLAVPAAPKK 765

Qy    128 EVLTAKT------VVTAPEAIR--------GTRVQGTWYQKTVQLPAGTKYVAFRHFG 171
           |||  |       ||| |||         |:  :|||| || |   | |: ||| :|
Db    766 AVAQPKTAVKAYTVTVKPQTTQTVSKIAQVKPNNTGIRASVYEKTAK--NGAKY-ADRTFY 822

Qy    172 CT-----------------DFFWINLDDVEIKANGKRADFTE--TFESSTHGEAPAE 209
           ||                 |  ::|: || ::| :  : ||     |:|  ||||
Db    823 VTKERAHGNETYVLLNNTSHNIPLGWFNVKDLNVQNLGKEVKTTQKYTVNKSNNGLSMVP 882

Qy    210 WTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWN-----GMALNPDNYLISKDV 263
           | ||| |  |    | |: :| ||  |   |:|||| | |      | || |  ||  |
Db    883 WGT------KNQVILTGNNI----AQGTFNATKQVSVGKDVYLYGTINNRTGWVNAKDL 931

Qy    264 TGATKVK--------YYAVNDGFPGDHYAVMISKTGT--------NAGDFTVFEETPN 307
           || |:||          :  | |  | |:||  | |           || ||| ||:|
Db    932 TAPTAVKPTTSAAKDYNYTYVIKNG-NGYYYVTPNSDTAKYSLKAFNEQPFAVVKEQVIN 990

Qy    308 G-------INKGGARFGLSTEADGAK--------PQSVWIERTVDLPAGTKY 344
           |        :||        ||| |:         || | |:||      :|
Db    991 GQTWYYGKLSNGKLAWIKST--DLAKELIKYNQTGMTLNQVAQIQAGLQY 1038


RESULT 6
OGP_MOUSE
ID   OGP_MOUSE     STANDARD;      PRT;   721 AA.
AC   Q62010;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Oviduct-specific glycoprotein precursor (Oviductal glycoprotein)
DE   (Oviductin) (Estrogen-dependent oviduct protein).
GN   OVGP1 OR OGP OR CHIT5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR; TISSUE=Oviduct;
RX   MEDLINE=96115001; PubMed=7492680;
RA   Sendai Y., Komiya H., Suzuki K., Onuma T., Kikuchi M., Hoshi H.,
RA   Araki Y.;
RT   "Molecular cloning and characterization of a mouse oviduct-specific
RT   glycoprotein.";
RL   Biol. Reprod. 53:285-294(1995).
CC   -!- FUNCTION: Binds to oocyte zona pellucida in vivo. May play a role
CC       in the fertilization process and/or early embryonic development.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- TISSUE SPECIFICITY: Epithelial cells of the oviduct.
CC   -!- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.
```

```
CC   -------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------
DR   EMBL; D32137; BAA06863.1; -.
DR   MGD; MGI:106661; Ovgp1.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR001579; Glyco_hydro_18AS.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   ProDom; PD000471; Glyco_hydro_18; 1.
DR   SMART; SM00636; Glyco_18; 1.
DR   PROSITE; PS01095; CHITINASE_18; FALSE_NEG.
KW   Glycoprotein; Fertilization; Repeat; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    721
FT   DOMAIN      486    632       OVIDUCT-SPECIFIC GLYCOPROTEIN.
FT                                [TAP]-G-[IV].
FT                                21 X 7 AA TANDEM REPEATS OF S-K-T-T-
FT   CARBOHYD    402    402       N-LINKED (GlcNAc...)    (POTENTIAL).
FT   CARBOHYD    442    442       N-LINKED (GlcNAc...)    (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GlcNAc...)    (POTENTIAL).
SQ   SEQUENCE   721 AA;  78807 MW;  37246C8F01665652 CRC64;

Query Match         5.1%;
Best Local Similarity   18.9%;   Pred. No. 0.28;
Matches  103;   Conservative   55;   Mismatches  185;   Indels  202;   Gaps   22;

Qy    51  FAGHNSAICA-----SSASYINF---------------------------------    68
          |||||       ::::|
Db   222  FTGHNSPLFSLPEDSKSSAYAMNYWRKLGTPADKLIMGFPTYGRNFYLLKESKNGLQTAS  281

Qy    69  EGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEE  128
          |
Db

Qy   282  MGPASPGKY----------------TKQAGFLAYYEVCSFVQRAKKHWIDYQ-------  317
          ::
Db   282  TKQAGFLAYYEVCSFVQRAKKHWIDYQ--------------------------------  317

Qy   129  VLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV----  183
             ||||         :||| ||    |||
Db   318  --YVPYAFKGKEWLG--YDDTISFSYKAMVKREIFGGAMVWTLDMDDVRGTFC       367

Qy   184  --------------------------------EIKANGKRAD---FTETFPSST-----  202
                                          ||||||||       ::||
Db   368  GNGPFPLVHILNELLVQTESNSTPLPQFWFTSSVNASGPGSENTALTEVLTTDTIKILPP  427

Qy   203  HGEAPA---------EWTTIDADGDGQGWLCLSSGQLGWLTAHGGT--NVVAS------  244
          ||           ::  |:|:
Db   428  GGEAMTTEVHRRYENMTTVPSDGS--VTPGGTASPRKHAVTPENNTMAAEAKTMST      481

Qy   245  --FSWNGMALNPDNYLISKDVTGATKV-----KYYYAVNDGFPGDHYAVM-ISKTGTNA  295
                     :|:|                    :|
Db   482  LDFFSKTTTGVSKTTTGISKTTTGVSKTTTGVSKATAGISKTIPEISKATAGVSKTTTGV  541

Qy   296  GDFTVVFEETPNGINKGGARFGLSTEADGAKPQSVWIER-TVDLPAGTKYVAFRHYNCSD  354
                     :| :
Db   542  SKTTTGISKTITGVSK--TTTGISKTTTGVSKTTTGVSKTTTGVSKTTTGISKTTTGIS-  598

Qy   355  LNYILLDDIQFTMGGSPTPTDYTVYRDGTKIKEGLTETT-----------FEEDGVAT  403
                        ::|:|
Db   599  ------------QTTTGISKTTTDISKTT-TGISKTTPGISKTTPGMTVIVQTQANEATTAT  648

Qy   404  GNHEYCVEVK-----YTAGVSPKE--------CVNNVTVDPVQFNPVQNLTGSAVGQK  447
          :|
Db   649  MDHQSVTPTEMDTTLFYLKMTPSEKETSRKKTMWLEKATVSPREMSATPN---------GQS  702

Qy   448  VTLKW  452
          ||||
Db   703  KTLKW  707
```

```
SLAP_LACBR
ID   SLAP_LACBR   STANDARD;   PRT;   465 AA.
AC   Q05044;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   S-layer protein precursor (Surface layer protein).
OS   Lactobacillus brevis.
OC   Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=1580;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=ATCC 8287 / DSM 20556;
RX   MEDLINE=93054358; PubMed=1429463;
RA   Vidgren G., Palva I., Pakkanen R., Lounatmaa K., Palva A.;
RT   "S-layer protein gene of Lactobacillus brevis: cloning by polymerase
RT   chain reaction and determination of the nucleotide sequence.";
RL   J. Bacteriol. 174:7419-7427(1992).
CC   -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC       of proteins which coat the surface of bacteria.
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a
CC       S-layer with tetragonal symmetry.
CC   -!- SIMILARITY: SOME, TO THE S-LAYER PROTEIN OF L.ACIDOPHILUS.
CC   -------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------
DR   EMBL; Z14250; CAA78618.1; -.
DR   PIR; A47023; A47023.
KW   Signal; Glycoprotein; Cell wall; S-layer.
FT   SIGNAL        1     30
FT   CHAIN        31    465       S-LAYER PROTEIN.
SQ   SEQUENCE   465 AA;  48159 MW;  2BE240339ZE65A2C CRC64;

Query Match         4.9%;
Best Local Similarity   19.8%;   Pred. No. 0.32;
Matches   99;   Conservative   48;   Mismatches  201;   Indels  153;   Gaps   20;

Qy    13  TTTLSSFEN-GIPASWKTIDA---------DGDGNNWTTTPPPGGTSFAGHNSAICASSASYIN   67
          ||   :|:   ||
Db    26  TTASAKSYATAGAYSTLKTDAATRNVEATGTNALYTKP--GTV---KGAKVVASKATMAK   80

Qy    68  FEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLE  127
          :::   :  |:|      :::|
Db    81  LASSKKSADYFRAYGVKTTNRGSVYRVTMDGKYRGYVYG----GKSDTAFAGGI--    132

Qy   128  EVLTAKTVVTAPEAIRGTRVQGTWYQKTVQ-----LPAGTKYVAFRHFGCTDFFWINLDD  182
          :|:                          |:|
Db   133  ----KSAETTKADMPARTTGFYLTDTSKNTLWTAPKYTQYKASK----------    173

Qy   183  VEIKANGKRADFTETFPSSTHGEAPAEWTTIDADGDG-QGWLCLSSGQLGWLTAHGGTNV  241
             :  :|     |:     |    ::  ||    :  |||:|  |:| |
Db   174  VSLYGVAKDTKFTVDQAATKTREGSLYYHVTATNGSGISGWIYAGK--GFSTTATGTQV  230

Qy   242  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  301
          :|                                 :|       :|
Db   231  LGGLSTDK--------SVTATNDNSVKIVYRTTD------GTQVGSNTWV  266

Qy   302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYV------AFRHYNCS  353
             |          :  ||:|     |:
Db   267  ---------------------TSTDGTKAGSKVSDKAADQTALEAYINANKPSGYTVNPNAA  308

Qy   354  DLNY-----ILLDDIQFTMGGSPTPTDYT-----------YTVYY  381
          |:|           |:        :
Db   309  DATYGNTVVATVSQAATSKVALKVSGTPVTTALTTADANDKVAANDTTANGSSVAGSTVY  368

Qy   382  RDGTKIKEGLTETTFE--------EDGVATGNHEYCVEV-----KYTAGVSPK  421

RESULT 7
```

```
Db    369  AAGTKLAQLTDLTGEKGQVVLTAIDTDLEDATFTGTTTYYSDLGKAYHYTYTNKDSA 428
                ||||:  |:    |||  ::       |: |:       |  ||     ||
Qy    422  ECVNVTVDPVQFNPVQNLTGS 442
                |::||    :|:|||
Db    429  ASSNAS--TQFG--SNVTGT 444
```

RESULT 8
```
GUXB_CELFI
ID   GUXB_CELFI     STANDARD;     PRT;  1090 AA.
AC   P50899;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Exoglucanase B precursor (EC 3.2.1.91) (Exocellobiohydrolase B)
DE   (1,4-beta-cellobiohydrolase B) (CBP120).
GN   CBHB OR CENE.
OS   Cellulomonas fimi.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Micrococcineae; Cellulomonadaceae; Cellulomonas.
OX   NCBI_TaxID=1708;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
RC   STRAIN=ATCC 484;
RX   MEDLINE=96003898; PubMed=7575482;
RA   Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT   "Cellobiohydrolase B, a second exo-cellobiohydrolase from the
RT   cellulolytic bacterium Cellulomonas fimi.";
RL   Biochem. J. 311:67-74(1995).
RN   [2]
RP   SEQUENCE OF 54-75.
RX   MEDLINE=93209933; PubMed=8458833;
RA   Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT   "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RT   D (CenD), a family A beta-1,4-glucanase.";
RL   J. Bacteriol. 175:1910-1918(1993).
RN   [3]
RP   SEQUENCE OF 54-78.
RX   MEDLINE=94197708; PubMed=8147863;
RA   Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.,
RA   Warren R.A.J., Miller R.C. Jr.;
RT   "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi
RT   CenE, a member of a new family of beta-1,4-glucanases.";
RL   Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
CC   -!- FUNCTION: Hydrolyze cellohexaose to a mixture of cellotetraose,
CC       cellotriose and cellobiose, with only a trace of glucose. It
CC       hydrolyzed cellopentaose to cellotriose and cellobiose, and
CC       cellotetraose to cellobiose, but it did not hydrolyze cellotriose.
CC       Has also weak endoglucanase activity. Hydrolyzes glucosidic bonds
CC       with inversion of anomeric configuration.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC   -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC       domain.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL
CC       HYDROLASES).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L38827; AAB00822.1; -.
DR   PIR; S59077; S59077.
DR   HSSP; P07986; 1EXG.
DR   InterPro; IPR001919; Bac_celose-bind.
DR   InterPro; IPR008965; Cellul_bind.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR000556; Glyco_hydro_48.
DR   InterPro; IPR008928; Glyco_trans_6hp.
DR   Pfam; PF00553; CBM_2; 1.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF02011; Glyco_hydro_48; 1.
DR   PRINTS; PR00844; GLHYDRLASE48.
DR   ProDom; PD011903; CBD_II; 1.
DR   SMART; SM00637; CBD_II; 1.
DR   SMART; SM00060; FN3; 3.
KW   Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT   SIGNAL         1     33       POTENTIAL.
FT   PROPEP        34     53
FT   CHAIN         54   1090       EXOGLUCANASE B
FT   DOMAIN        54    699       CATALYTIC (BY SIMILARITY).
FT   DOMAIN       700    785       FIBRONECTIN TYPE-III 1
FT   DOMAIN       794    884       FIBRONECTIN TYPE-III 2.
FT   DOMAIN       891    978       FIBRONECTIN TYPE-III 3.
FT   DOMAIN       989   1090       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE     513            NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID     990    513       BY SIMILARITY.
SQ   SEQUENCE  1090 AA;  114829 MW;  046BB9D956F2F399 CRC64;

Query Match             4.9%;   Score 121;  DB 1; Length 1090;
Best Local Similarity  22.3%;   Pred. No. 1.1;
Matches  106; Conservative  53; Mismatches 197; Indels 120; Gaps  26;

Qy     26  ASWKT---IDADGDGNNWTTTPPPGG------TSFAGHNSAICASSASYINFEGPQNPD  75
               ||||     ||: |  ||  |:|  |        |:|:|  |  |  : : : :|
Db    515  ASWKVPSELKWTGKPDTWNAAAPTGNPGLITVEVTSY-GQDVGVAADTARALLFYAAKSGD 573

Qy     76  NYLVTPELSLPNGGTLTFWVCAQDANYASBHYAVVASSTGNDASNFANALLEEVLTAKTV 135
               |:|   |  |         ||: |: :| ||  |  : :| |   | ||  :| :||
Db    574  TASRDKAKALLD----AIWANNQDP---------LGVSAVETRGDYKRFDDTYVAN----GDG 619

Qy    136  VTAPEAIRGTRVQG----------------------TWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDD 182
               |||  |  |                          ::|:|  :|| :| : |||   :
Db    620  IYIPSGWTGTMPNGDVIKPGVSFLDIRSFYKKD---PNWSKVQTFLDGGAEPQPRYHRFW 676

Qy    183  VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVV 242
               |:|  | :| ||      ::  :|     ||| | | |  | |: :| |  : :||
Db    677  AQTAVAGALADYARLFDDGT--------------------TTPDTTAP---TVPTG---LQAGVVTSTE 719

Qy    243  ASFSWNGMALNPDNYLISKDV-TGATKVKYYAVNDGPFGDHYAVMISKTGTNAGD---F 298
               |:|: || |    |:| ||| |     ||:  |  ||: | :  :|| ||||
Db    720  ATISWT--ASTDDTRVTGYDVYRGATKV----------------GTATTTSFTDTGLTASTAYAY 766

Qy    299  TVFEETNGINKGGARFGLSTEA---DGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 355
               ||  | :| :|| : | ::|:   |  |||  : |:  :|  :|| :| :|:  |
Db    767  TVRAFDAAGNVSAPSAALTVTTKATPSDTTAP-------SVPAITS------SSSTA 810

Qy    356  NYILLDIQFTMGGSPTPTDYT-------YTVYRDGTKIKEGLTETFEEDGV-ATGNHE 407
               | :|:  |||||:| |  ||:       |||||||: ||:| ||:: |  |||||
Db    811  NSV------TIGWSAS-TDNAGGSGLAGYDVYRGATVRAQ-TTALTFTDTGLTASTAYE 861

Qy    408  YCVEVKYTAG--VSPKECVNVTV-----------------DPVQFNPVQNLTGSAVGQ-KVTLKWDA 454
               | :||:|: |  :|||| |||| |                 |||:|||||:|||:|| :|: |||
Db    862  YTVRARDVAGNVSAPSTAVSVTTKSDTTPDTTAPSVPAGLAAMTVTETSVALTWNA 917
```

RESULT 9
```
C5AA_BACUD
ID   C5AA_BACUD     STANDARD;     PRT;  1385 AA.
AC   Q45760;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Pesticidal crystal protein cry5Aa (Insecticidal delta-endotoxin
DE   CryVA(a)) (Crystaline entomocidal protoxin) (152 kDa crystal protein).
GN   CRY5AA OR CRYVA(A) OR CRYVA.
OS   Bacillus thuringiensis (subsp. darmstadiensis).
```

```
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=132264;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL B-18243 / PS17;
RA   Sick A.J., Schwab G.E., Payne J.M.;
RT   "Genes encoding nematode-active toxins cloned from Bacillus
RT   thuringiensis isolate PS17.";
RL   Patent number US5281530, 25-JAN-1994.
CC   -!- FUNCTION: ENDOTOXIN WITH NEMATICIDAL ACTIVITY.
CC   -!- DEVELOPMENTAL STAGE: The crystal protein is produced during
CC       sporulation and is accumulated both as an inclusion and as part of
CC       the spore coat.
CC   -!- MISCELLANEOUS: Toxic segment of the protein is located in the N-
CC       terminus.
CC   -!- SIMILARITY: Belongs to the delta endotoxin family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L07025; AAA67694.1; -.
DR   PIR; T18213; T18213.
DR   InterPro; IPR005638; endotoxin_C.
DR   InterPro; IPR005639; endotoxin_N.
DR   InterPro; IPR008979; Gal_bind_like.
DR   Pfam; PF03944; endotoxin_C; 1.
DR   Pfam; PF03945; endotoxin_N; 1.
KW   Toxin; Sporulation.
SQ   SEQUENCE   1385 AA;   152439 MW;   E29FF11FC799DE95 CRC64;

Query Match             4.8%;
Best Local Similarity  20.1%;
Matches  120;  Conservative  66;  Mismatches  192;  Indels  219;  Gaps  30;

                       Score 118.5;   DB 1;   Length 1385;
                       Pred. No. 2.2;

Qy      5 PNPNPNPGTTTLSESFPENGI---PASWKTIDADGDGNNWTTTPPPG---GTSF------ 51
            :  :  :.|| |     |           |  | ||  |    |         ||
Db    413 PAVDPQAGPNVSIDSSNPIIQINMDTWKTPPQGASG--WNTNLMRGSVSGLSFLQRDGT 470

Qy     52 -----AGHNSAICASSA---SYINFEGPQNPDNY------LVTP-ELSLPN--- 87
               ||| |  | | |      ||| |  |          |  | | | :
Db    471 RLSAGMGGGFADTIYSLPATHYLSYLYGTPYQTSDNYSGHVGALIVGVSTPQEATLPNIIG 530

Qy     88 ------------------------GGTLT-FWVCAQDA------ 112
                                      |||  : | : : |      :
Db    531 QPDEQGNVSTMGFPFPEKASYGGTVVKEWLNGANAMKLSPGGSIGIPITNVTSGEYQIRCR 590

Qy    113 STGNDASNF--------ANALLEEVLTAKTVVTAPEAIRGTRVQG---------TWYQ 153
          |  | :     :     |     |:    |   |     |             ||:
Db    591 YASNDNTNVFFNVDTGGANPIFQQINFASTV------DNNTGVQGANGVVVVKSIATTDN 644

Qy    154 KTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGKRADFTEIFESSTHGEAPAEWTTI 213
           ::|||    | |:  |   |   || |:|   | |  | |  :     |  |
Db    645 SFTEIPAKTINVHLTNQGSSDVF--LDRIEFIPFSLPLIYHGSYNTSS-GADDVLW--- 697

Qy    214 DADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWNGMALNPDNYLISK------DVTGA 266
          : | :  | |   |  :    |  |    |: |: :|     :| | |       | |
Db    698 ----------SSSNMNYY------DIIVNGQANSSSIASSMHILNKGKVIKTIDIPGH 739

Qy    267 TKVKY-YYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGAARFGLSTEADGA 325
            :  : :    | :|   |  |    |   | |     |  | |  | |  |
Db    740 SETFFATFPVPEGF--NEVRILAGLPEVSGNITVQSNNPPQPSNNGGGDGGNGGGDGG 796

Qy    326 KPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPT---------- 374
          :                ::  ::  ||  :   : || | ||  |
Db    797 Q---------------------YN------FSLSGSDHTTIYHGKLETGIH 820

Qy    375 ---DYTYT--------VVRDGTKIKEGLITETTF----EED--------GVATGNHE 407
             ::|||         :    | ||  |  |        |          |  |||
```

DR   PRINTS; PR00733; GLHYDRLASE6.
DR   ProDom; PD003733; Glyco_hydro_6; 2.
DR   SMART; SM00637; CBD_II; 1.
DR   SMART; SM00060; FN3; 3.
DR   PROSITE; PS00561; CBD_BACTERIAL; 1.
DR   PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR   PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT   SIGNAL   1   40
FT   CHAIN   41   872   EXOGLUCANASE A.
FT   DOMAIN   41   477   CATALYTIC.
FT   DOMAIN   478   563   FIBRONECTIN TYPE-III 1.
FT   DOMAIN   573   664   FIBRONECTIN TYPE-III 2.
FT   DOMAIN   673   768   FIBRONECTIN TYPE-III 3.
FT   DOMAIN   769   872   CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE   188   PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   410   NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   140   202   BY SIMILARITY.
FT   DISULFID   374   428   BY SIMILARITY.
FT   DISULFID   770   869   BY SIMILARITY.
SQ   SEQUENCE   872 AA; 89300 MW; 7883B407F995533B CRC64;

Query Match
Best Local Similarity   4.6%;   Score 114; DB 1; Length 872;
Matches 92; Conservative 29; Mismatches 146; Indels 112; Gaps 19;

Qy   12 GTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYINFEGP   71
Db   491 GTTTAT-----SVPLSW---TASTDNVAVTGYDVYRGTTLVGTTAA------------- 528

Qy   72 QNPDNYLVTPELSLPNGGTLTFWVCAQDA--NYASEHYAVYASSTGNDASNFANALLEEV 129
Db   529 ---TSYTVT---GLTPATAYSFTVRAKDAAGNVSAASAAAAATTQSGTVTDTTAPSVPAG 582

Qy   130 LTAKTVVTAP------------------------EAIRGTRVQGTW----YQKTVQLPAGTKYV 165
Db   583 LIAGTTTTTTVPLSWTASTDNAGGSGVAGYEVLRGTTVVGTTTATSYIVT-GLTAGTTY- 640

Qy   166 AFRHFGCTDFFWINLDDVEIKANGKRADFTETFE-----------------SSTHGEA 206
Db   641 SFSVRAKDVAGNTSAASAAVSATTQTGTVTVVDTTAPSVPTGLTAGTTTTSSV 691

Qy   207 PAEWTTIDADGDGDQGWLCLSSGQLGWLTAHGGTNV--VASFSWNGMALNPDN----YLIS 260
Db   692 PLTWT---ASTDNAG-------GSGVAGYEVFNGTTRVATVTSTSYTVTGLAADTAYSFTVKA 744

Qy   261 KDVTG-------ATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINK-G 312
Db   745 KDVAGNVSAASAAVSARTQAATSGG-----CTVKYSASSWNTG-FTGTVEVKNNGTAALN 798

Qy   313 GARFGLSTEADGAKPQSVW 331
Db   799 GWTLGFSF-ADGQKVSQGW 816

RESULT 11
YEEJ_ECO57
ID   YEEJ_ECO57   STANDARD;   PRT;   2660 AA.
AC   Q8XBV7; Q8X2B9; Q8X2C0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein yeeJ
GN   Z3135 OR ECS2775/ECS2776.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533 (2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- SIMILARITY: Contains 16 Big-1 domains.
CC   -!- SIMILARITY: Belongs to the intimin/invasin family.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC   frameshift in position 1315.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AE005423; AAG57041.1; ALT_FRAME.
DR   EMBL; AP002559; BAB36198.1; ALT_FRAME.
DR   EMBL; AP002555; BAB36199.1; Big_1.
DR   InterPro; IPR003344; Big_1.
DR   InterPro; IPR003535; Intimin.
DR   InterPro; IPR008964; Invasin_intimin.
DR   InterPro; IPR000601; PKD.
DR   Pfam; PF02369; Big_1, 16.
DR   PRINTS; PR01369; INTIMIN.
DR   SMART; SM00634; BID_1; 16.
DR   SMART; SM00089; PKD; 8.
KW   Hypothetical protein; Repeat; Complete proteome.
FT   DOMAIN   738   834   BIG-1 1.
FT   DOMAIN   840   929   BIG-1 2.
FT   DOMAIN   931   1033   BIG-1 3.
FT   DOMAIN   1042   1132   BIG-1 4.
FT   DOMAIN   1134   1236   BIG-1 5.
FT   DOMAIN   1245   1335   BIG-1 6.
FT   DOMAIN   1337   1439   BIG-1 7.
FT   DOMAIN   1448   1539   BIG-1 8.
FT   DOMAIN   1548   1652   BIG-1 9.
FT   DOMAIN   1653   1750   BIG-1 10.
FT   DOMAIN   1751   1855   BIG-1 11.
FT   DOMAIN   1856   1957   BIG-1 12.
FT   DOMAIN   1963   2056   BIG-1 13.
FT   DOMAIN   2065   2156   BIG-1 14.
FT   DOMAIN   2157   2252   BIG-1 15.
FT   DOMAIN   2254   2355   BIG-1 16.
SQ   SEQUENCE   2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match
Best Local Similarity   4.6%;   Score 113; DB 1; Length 2660;
Matches 104; Conservative 58; Mismatches 199; Indels 106; Gaps 24;

Qy   13 TTTLSESFE---NGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYINFE   69
Db   1470 TATVKDQFDNEVNNLPVTFSTA-----SSGLTLTPGESNTNESGIAQATLAGVAF----- 1519

Qy   70 GQNPDNVLVTPELSLPNGG--------TLTF--WVCAQDANYASEHYAVYASSTGNDASNF 121
Db   1520 GEQ------TVTASLANNGASDNKTVHFIGDTAAAKIIELTPVPDSIIAGTPQNSSGS- 1571

```
FT   ACT_SITE        481    481                   PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE        547    547                   NUCLEOPHILE (BY SIMILARITY).
SQ   SEQUENCE       1034 AA;  118673 MW;  3B644C9A649415E9 CRC64;

Query Match        4.5%;       Score 112.5;  DB 1;  Length 1034;
Best Local Similarity  19.9%;     Pred. No. 3.9;
Matches 114;  Conservative 70;  Mismatches 169;  Indels 221;  Gaps  35;

Qy     122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVVQLPAGTKYVAFRHFGCTDFFWINLD  181
                 |:||           :| :||  ||   :|    ::      :
Db    1572  ------------VITATVVDNNGFPVKGVTVVNFTSNAATAEMTNGGQAVTNEQGKATVTYTNTRS 1624

Qy     182  DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQ-GWLCLSSGQLGWLTAHGGTN  240
                 :| :|| :  :|        ||||||||||||||    |||||||||||||:||:|
Db    1625  SIE---SGARPD---TVEASLENGSSTLSTSINVNADASTAHLTLLQALFDIVSAGDTTN 1678

Qy     241  V---VASFSWNG----MAINPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISK  290
                :  ::::|        |||     ::|| ||     | :::
Db    1679  LYIEVKDNYGNGVPQQEVTLSVSP-----SEGVTPSNNAIY---TTNHDGNFYA---SF  1726

Qy     291  TGTNAGDFTVV------FEETPNGI-NKGGARFGLSTEADGAKPQSVWIERTVDLPAG  341
                ||        |||     |||     ||||
Db    1727  TATKAGVYQVTATLENGDSMQQVTYVPNVANAEISLAASKDPV------IANNNDLTTL 1780

Qy     342  TKYVAFRHYNC---SDLNYILLDDIQ--FTMGGSPTPTDYITTVYRDGTKI---KEGLTE  393
                |||              |||||       :  |  ::||  |
Db    1781  TATVADTEGNAIANSEVTFTLPEDVRANFTLG----------DGGKVVTDTEGKAK  1826

Qy     394  TTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLT 440
                |:            :|           :|:
Db    1827  VTLK--GTKAGAH--TVTASMAGGKSEQLVVNFIADTLITAQVNLNVT 1869
```

RESULT 12
```
BGAL_BACME
ID   BGAL_BACME     STANDARD;      PRT;   1034 AA.
AC   O52847;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Beta-galactosidase (EC 3.2.1.23) (Lactase).
GN   BGAM.
OS   Bacillus megaterium.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1404;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 319;
RA   Strey J.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC       galactose residues in beta-D-galactosides.
CC   -!- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.
```

```
DR   EMBL; AJ000733; CAA04267.1; -.
DR   PIR; T30574; T30574.
DR   HSSP; P00722; 1BGL.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR006101; Glyco_hydro_2.
DR   InterPro; IPR006102; Glyco_hydro_2ig.
DR   InterPro; IPR006104; Glyco_hydro_2SB.
DR   InterPro; IPR006103; Glyco_hydro_2TIM.
DR   InterPro; IPR004200; Glyco_hydro_42C.
DR   InterPro; IPR004199; Glyco_hydro_42N.
DR   Pfam; PF02930; Bgal_small_C; 1.
DR   Pfam; PF02929; Bgal_small_N; 1.
DR   Pfam; PF00703; Glyco_hydro_2; 1.
DR   Pfam; PF02836; Glyco_hydro_2_C; 1.
DR   Pfam; PF02837; Glyco_hydro_2_N; 1.
DR   PRINTS; PR00132; GLHYDRLASE2.
DR   PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR   PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KW   Hydrolase; Glycosidase.
```

RESULT 13
```
ALS1_CANAL
ID   ALS1_CANAL     STANDARD;      PRT;   1260 AA.
AC   P46590;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Agglutinin-like protein 1 precursor.
GN   ALS1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11651 / B792;
RX   MEDLINE=95272392; PubMed=7752895;
RA   Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RT   "Candida albicans ALS1: domains related to a Saccharomyces cerevisiae
RT   sexual agglutinin separated by a repeating motif.";
RL   Mol. Microbiol. 15:39-54(1995).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-glycosylated and O-glycosylated (Potential).
CC   -!- SIMILARITY: TO YEAST SAG1.
```

```
Db    566 ------------------PGGTDSVIREPPNHTVTTTEYW----------------- SQSYA 593
                            |||:|||||||||||||||||
Qy    200 SSTHGEAPAEWTTIDADGDQGWLCLSSQLGWLTAHGGTNVASFSWNGMALNPDNYLI 259

Db    594 TTTTVTAPP-------------------------------------GGTDTV------IIREPPNHTV 618
          |:|:||                                        |||||:
Qy    594 TTTTVTAPP-----------------------------------GGTDTV------IIREPPNHTV

Qy    260 SKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPN 307
                                          |||:::|:
Db    619 ------------SQSFATTTTVTGPPSGTDVIIREPPN 651
                      :|:: |
          TTTEYW-----SQSFATTTTVTGPPSGTDVIIREPPN
```

RESULT 14
```
ADP1_MYCGA
ID   ADP1_MYCGA    STANDARD;      PRT;   1122 AA.
AC   Q49379; Q49437; Q53351;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Adhesin P1 precursor (Cytadhesin P1) (Attachment protein).
GN   GAPA OR MGC1.
OS   Mycoplasma gallisepticum.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2096;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S6;
RX   MEDLINE=96201559; PubMed=8613358;
RA   Keeler C.L. Jr., Hnatow L.L., Whetzel P.L., Dohms J.E.;
RT   "Cloning and characterization of a putative cytadhesin gene (mgc1)
RT   from Mycoplasma gallisepticum.";
RL   Infect. Immun. 64:1541-1547(1996).
RN   [2]
RP   SEQUENCE OF 1-12 FROM N.A.
RC   STRAIN=S6;
RA   Hnatow L.L., Keeler C.L. Jr., Tessmer L., Dohms J.E.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 378-570 FROM N.A.
RC   STRAIN=S6;
RX   MEDLINE=93371270; PubMed=8363503;
RA   Dohms J.E., Hnatow L.L., Whetzel P., Morgan R., Keeler C.L. Jr.;
RT   "Identification of the putative cytadhesin gene of Mycoplasma
RT   gallisepticum and its use as a DNA probe.";
RL   Avian Dis. 37:380-388(1993).
RN   [4]
RP   SEQUENCE OF 159-1122 FROM N.A.
RC   STRAIN=S6;
RA   Goh M.S., Geary S.J.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Could be involved in cytadherence.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the adhesin P1 family.
CC   -!- CAUTION: Ref.4 sequence differs from that shown due to a
CC       frameshift in position 159 to 213.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U34842; AAB02987.1; -.
DR   EMBL; U44804; AAC83385.1; ALT_FRAME.
DR   PIR; T18346; T18346.
KW   Cytadherence; Signal; Transmembrane.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1122       ADHESIN P1.
FT   TRANSMEM    997   1021       POTENTIAL.
FT   DOMAIN     1002   1009       POLY-ILE.
FT   CONFLICT    313    314       DM -> IW (IN REF. 4).
FT   CONFLICT    382    382       Y -> F (IN REF. 3).
```

```
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L25902; AAC41649.2; -.
DR   InterPro; IPR008440; Candida ALS.
DR   Pfam; PF05792; Candida ALS; 1.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1260       AGGLUTININ-LIKE PROTEIN 1.
FT   DOMAIN      433    792       10 X 36 AA TANDEM REPEATS.
FT   REPEAT      433    468       1-1.
FT   REPEAT      469    504       1-2.
FT   REPEAT      505    540       1-3.
FT   REPEAT      541    576       1-4.
FT   REPEAT      577    612       1-5.
FT   REPEAT      613    648       1-6.
FT   REPEAT      649    684       1-7.
FT   REPEAT      685    720       1-8.
FT   REPEAT      721    756       1-9.
FT   REPEAT      757    792       1-10.
FT   DOMAIN      983   1152       2 X 26 AA APPROXIMATE REPEATS.
FT   REPEAT      983   1043       2-1.
FT   REPEAT     1092   1152       2-2.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      450    455       POLY-THR.
FT   DOMAIN      486    491       POLY-THR.
FT   DOMAIN      522    527       POLY-THR.
FT   DOMAIN      558    563       POLY-THR.
FT   DOMAIN      594    599       POLY-THR.
FT   DOMAIN      630    635       POLY-THR.
FT   DOMAIN      666    671       POLY-THR.
FT   DOMAIN      702    707       POLY-THR.
FT   DOMAIN      738    743       POLY-THR.
FT   DOMAIN      774    779       POLY-THR.
FT   DOMAIN      874    877       POLY-SER.
FT   CARBOHYD    471    471       N-LINKED (GLCNAC...).  (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC...).  (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC...).  (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC...).  (POTENTIAL).
FT   CARBOHYD    723    723       N-LINKED (GLCNAC...).  (POTENTIAL).
FT   CARBOHYD    820    820       N-LINKED (GLCNAC...).  (POTENTIAL).
FT   CARBOHYD    886    886       N-LINKED (GLCNAC...).  (POTENTIAL).
FT   CARBOHYD    918    918       N-LINKED (GLCNAC...).  (POTENTIAL).
FT   CARBOHYD    973    973       N-LINKED (GLCNAC...).  (POTENTIAL).
FT   CARBOHYD   1045   1045       N-LINKED (GLCNAC...).  (POTENTIAL).
FT   CARBOHYD   1068   1068       N-LINKED (GLCNAC...).  (POTENTIAL).
SQ   SEQUENCE  1260 AA;  132641 MW;  763D1063A2354C24 CRC64;
```

```
Query Match       4.5%;  Score 112.5;  DB 1;  Length 1260;
Best Local Similarity  19.5%;  Pred. No. 5;  Indels 173;  Gaps 16;
Matches   68;  Conservative  30;  Mismatches  77;

Qy      5 PNPNPNPGTTTL-SESFENGIPASWKTIDADGDGNN--------------------W------ 40
          |:|||  |||||  |:|:||      |:|
Db    432 PSPNPTVSTTEYWSQSF------ATTTTVTAPPGGTDFVIIREPPNHTVTTTEYWSQSFAT 486

Qy     41 --TTTPPGGTSFAGHNSAICASSASYINFEGPQNP-------DNYLIVTPELSLPNGG 89
            ||| |||
Db    487 TTTTVTAPPGGT----DSVI------IREPPNPTVTTTEYWSQSFATTTTVTAPPGG 532

Qy     90 TLTFWV-----CAQDANYASEHYAVVASSTGNDASNFANALLEEVLTAKTVTAPEAIR 143
          |:: |
Db    533 TDSVIREPPNPTVTTTEYWSQSYA------------------TTTTVTAP---- 565

Qy    144 GTRVQGTWYQKTVQLPAGTKYVAFR---HFGCTDFFWINLDDVEIKANGKKRADFTETFE 199
                                         :|| ::
```

```
FT   CONFLICT     501    501      I -> T (IN REF. 3).
FT   CONFLICT     568    568      R -> G (IN REF. 3).
FT   CONFLICT     570    570      T -> A (IN REF. 3).
FT   CONFLICT     693    695      DIL -> VYT (IN REF. 4).
SQ   SEQUENCE    1122 AA;  121305 MW;  15SC34DA2D6C3C65 CRC64;

Query Match             4.5%;
Best Local Similarity  19.0%;    Score 110.5;  DB 1;  Length 1122;
Matches    94;  Conservative   77;  Mismatches  193;  Indels  131;  Gaps   26;

Qy    42 TTPPPGGTSFAGHNSAICASSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDAN 101
            |:|:.:                    ||:|:                 |:|||
Db   437 TABAPGNTKTVGYPYGILLSA---ISFDATR------------------NGLALAAPALGQDVG 479

Qy   102 YASEHYA------------VYASST---GNDASNFAN------ALLEEVLTAKTV 135
            |     |          :|:.           |:|:|      ||
Db   480 Y---HFVPRLAVGGVSSPRGANGNIFLGSAITWGTNGGNFLDTKWHSPAVIEDAPTTFIT 536

Qy   136 VTAPEAIRGTRVGGTWYQKTVQLP--AGTKYVAFRHFGCTDFFWINLLDDVEIK---ANGK 190
            |  .:  |: .:          ||             |:|:  ::   |  .
Db   537 VNSSGVLQNS---GSQQSTSTPMPNSNGNESIPYRWTNSYDYNSVRFAALISKPAGGNTK 593

Qy   191 RAD--FTETFESSTHGEAPAEWTTID------ADGDGQWLCLSSGQLGWLTAHGGTNVV 242
            |||    : : :|:||             ||:  || |||
Db   594 QVESLFTTALKLDTLNSLPNKFTQENNIFFSYAMLDGRQWSLGTRKDSTWLT----TNTI 649

Qy   243 ASFSWN---------------GMALNPDNYLIS-----------KDV-------TGATKVKYYAV 275
            .:|:|                 :::                   :|:
Db   650 NNFTYNTQQQLASTAAGENANPRNILNALTTAKGFDRRDIGNVDILIYSNNINKFTYYYQV 709

Qy   276 ND----------GFPGDHYAVMISKTGTNAGDFTVVFEETPNGINK-GGARFGLSTEADG 324
                         :|  ||        |     |:|::   :|    |:|::|
Db   710 GGAITTWPEVQVNYKTSANITYNLTRTDFGSTTPATQDANTVSSKLNGAYLSSTGDQQG 769

Qy   325 AKPQSVWIERTVDLPAGTKYV--AFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYR 382
            ||::: .  |:|::  :|:|       :|:  :       :      ::|
Db   770 WYNGSIYVKKASFTPSSQGYTWQDFKGLTTTASNAVISN----WTKAG--------YSIRP 818

Qy   383 DGTKIKEGLTETTFEEDGVATGN----HEYCVE-----VKYTAGVSPKECVNVTVDPVQ 432
            |: |: : .| :                |:||      :|:::|
Db   819 DDDTV-FSVSKIPFEKEITAAVNVRSLDSYYVQLNGETSVNTVARVSPDSSA-LTLNPKR 876

Qy   433 F-NPVQNLTGSAVGQ 446
            | :|| .: :|:|
Db   877 ITNPLMN-RDNVIGQ 890
```

```
RESULT 15
PME_ASPAC
ID   PME_ASPAC     STANDARD;       PRT;   331 AA.
AC   Q12535;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Pectinesterase precursor (EC 3.1.1.11) (Pectin methylesterase) (PE).
GN   PME1.
OS   Aspergillus aculeatus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5053;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KSM 510;
RX   MEDLINE=97079238; PubMed=8920970;
RA   Christgau S., Kofod L.V., Halkier T., Andersen L.N., Hockauf M.,
RA   Dorreich K., Dalboege H., Kauppinen S.;
RT   "Pectin methyl esterase from Aspergillus aculeatus: expression
RT   cloning in yeast and characterization of the recombinant enzyme.";
RL   Biochem. J. 319:705-712(1996).
CC   -!- FUNCTION: Involved in maceration and soft-rotting of plant tissue.
CC   -!- CATALYTIC ACTIVITY: Pectin + N H(2)O = N methanol + pectate.
CC   -!- SIMILARITY: Belongs to the pectinesterase family.
```

```
DR   EMBL; U49378; AAB42153.1; -.
DR   InterPro; IPR000070; Pectinesterase.
DR   Pfam; PF01095; Pectinesterase; 1.
DR   PROSITE; PS00800; PECTINESTERASE_1; 1.
DR   PROSITE; PS00503; PECTINESTERASE_2; 1.
KW   Hydrolase; Aspartyl esterase; Cell wall; Signal.
FT   SIGNAL         1     17
FT   CHAIN         18    331       Pectinesterase.
FT   ACT_SITE     162    162       BY SIMILARITY.
FT   ACT_SITE     183    183       BY SIMILARITY.
SQ   SEQUENCE     331 AA;  35681 MW;  1F1C81BF1E32174F CRC64;

Query Match             4.4%;
Best Local Similarity  22.5%;    Score 109.5;  DB 1;  Length 331;
Matches    69;  Conservative   33;  Mismatches   99;  Indels  105;  Gaps   15;

Qy   108 AVTASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGT-KYVA 166
            |:.|:. .||                ||:|:||.|.:|
Db    27 AIVVAKSGGDYTTIGDAI---DALSTSTTDTQTIFIE---EGT-YDEQVLPAMTGKVII 79

Qy   167 FRHFGCTDFFWINL---DDVEIKANGKRADFTETFE------------SSTHGEAPAE 209
            :|:    | :||     || :|                         |::||:||
Db    80 YGQTENTDSYADNLVTIHAISYEDAGESDDLTATFRNKAVGSQVYNLNIANTCGQACHQ 139

Qy   210 WTTIDADGDGQGWL-CLSSGQLGWLTAHGGTNVV- 242
            |:| :| ||||| |  ::||:|:||
Db   140 ALALSAWADQQGYYGCNFTGYQDTLLAQTGNQLYINSYIEGAVDFIFGQHARAWFQNVDI 199

Qy   243 -----ASFSWNGMALNPDN--YLISKDVTGA------TKVKYYAVNDGFPGDHYAV 286
               :||::|:               ::
Db   200 RVVEGPTSASITANGRSSEFDTSYVINKSTVAAKEGDDVAEGTYYL---GRPWSEYA- 254

Qy   287 MISKTGTNAGDFTVVFEET--PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKY 344
            | :|:|                |:|: ::::|:|
Db   255 ----------RVVFQQTSMTNVINSLG--------------------WTEWSTSTP-NTEY 284

Qy   345 VAFRHY 350
            |:|:|:
Db   285 VTFGEY 290
```

Search completed: May 18, 2004, 11:43:34
Job time : 10.5342 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:36:20 ; Search time 33.3368 Seconds
(without alignments)
4315.838 Million cell updates/sec

Title: US-08-570-311-18
Perfect score: 2480
Sequence: 1 GTNPNPNPNPGTTTLSESF........QNLTGSAVGQKVTLKWDAPN 456

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:
```
 1: sp_archea:*
 2: sp_bacteria:*
 3: sp_fungi:*
 4: sp_human:*
 5: sp_invertebrate:*
 6: sp_mammal:*
 7: sp_mhc:*
 8: sp_organelle:*
 9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2094	84.4	1723	2	P72197	P72197 porphyromon
2	2082	84.0	1687	2	Q9R9B7	Q9R9B7 porphyromon
3	2082	84.0	1704	2	Q5I816	Q5I816 porphyromon
4	2079	83.8	1723	2	P72194	P72194 porphyromon
5	2071	83.5	1358	2	P96967	P96967 porphyromon
6	2070	83.5	1706	2	Q5I839	Q5I839 porphyromon
7	2059	83.0	1706	2	Q5I838	Q51838 porphyromon
8	2054.5	82.8	1223	2	Q92NB5	Q92NB5 porphyromon
9	2039.5	82.2	1097	2	P72196	P72196 porphyromon
10	2039.5	82.2	1732	2	O07442	O07442 porphyromon
11	2034.5	82.0	1732	2	Q5I817	Q5I817 porphyromon
12	2020.5	81.5	1732	2	O52050	O52050 porphyromon
13	682	27.5	925	2	Q9F4J0	Q9F4J0 porphyromon
14	561.5	22.6	312	2	Q9KIB3	Q9KIB3 porphyromon
15	233	9.4	293	2	Q9XBU9	Q9XBU9 porphyromon
16	152.5	6.1	1742	16	Q8P377	Q8P377 xanthomonas
17	144	5.8	3346	16	Q7WN54	Q7wn54 bordetella
18	139.5	5.6	1341	16	Q8UAU1	Q8uau1 agrobacteri
19	139	5.6	2215	16	Q7WBN0	Q7wbn0 bordetella
20	135	5.4	3988	17	Q8TPZ1	Q8tpz1 methanosarc
21	134.5	5.4	2768	16	Q8E9G6	Q8e9g6 shewanella
22	134	5.4	1541	16	Q7UQJ9	Q7uqj9 rhodopirell
23	134	5.4	1800	2	Q9L94B	Q9l94b pseudomonas
24	134	5.4	8682	16	Q88RG2	Q88rg2 pseudomonas
25	133.5	5.4	680	2	Q52644	Q52644 ruminococcu
26	133	5.4	2522	16	Q8EKA6	Q8eka6 shewanella
27	132.5	5.3	555	16	Q826W1	Q826w1 streptomyce
28	132.5	5.3	691	16	Q9RZS7	Q9rzs7 deinococcus
29	132.5	5.3	1115	4	Q9HD43	Q9hd43 homo sapien
30	131.5	5.3	1256	16	Q8NX96	Q8nx96 staphylococ
31	131.5	5.3	1357	17	Q8THC8	Q8thc8 methanosarc
32	131.5	5.3	2468	16	Q912M3	Q912m3 pseudomonas
33	131	5.3	2219	16	Q88WI9	Q88wi9 lactobacill
34	130.5	5.3	880	2	Q9RMB8	Q9rmb8 arthrobacte
35	130.5	5.3	955	17	Q8TQ91	Q8tq91 methanosarc
36	130	5.2	1386	17	Q8TI72	Q8ti72 methanosarc
37	129.5	5.2	756	9	Q858B6	Q858b6 enterobacte
38	129.5	5.2	1744	16	Q82YW8	Q82yw8 enterococcu
39	129	5.2	7716	16	Q7UWZ8	Q7uwz8 rhodopirell
40	128.5	5.2	1248	16	Q99V41	Q99v41 staphylococ
41	128.5	5.2	1255	5	Q7WTC6	Q7wtc6 staphylococ
42	128.5	5.2	2656	5	Q9GNU3	Q9gnu3 paracentrot
43	128	5.2	1672	16	Q8Y366	Q8y366 ralstonia s
44	128	5.2	4936	16	Q8YKJ3	Q8ykj3 anabaena sp
45	127.5	5.1	1684	2	Q03658	Q03658 unidentifie

ALIGNMENTS

```
RESULT 1
P72197
ID    P72197      PRELIMINARY;    PRT;   1723 AA.
AC    P72197;
DT    01-FEB-1997 (TrEMBLrel. 02, Created)
DT    01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE    Lys-gingipain.
GN    KGP.
OS    Porphyromonas gingivalis (Bacteroides gingivalis).
OC    Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC    Porphyromonadaceae; Porphyromonas.
OX    NCBI_TaxID=837;
RN    [1]
RP    SEQUENCE FROM N.A.
RA    Payloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA    Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT    "Molecular cloning and characterization of Porphyromonas gingivalis
RT    Lys-gingipain.";
RL    Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; U54691; AAA99810.1; -.
DR    MEROPS; C25.002; -.
DR    GO; GO:0005524; F:ATP binding; IEA.
DR    GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR    GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR    GO; GO:0006310; P:DNA recombination; IEA.
DR    GO; GO:0006281; P:DNA repair; IEA.
DR    GO; GO:0006260; P:DNA replication; IEA.
DR    GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR    InterPro; IPR0000977; DNA_ligase.
DR    InterPro; IPR001769; Peptidase_C25.
DR    InterPro; IPR005536; Peptidase_C25_C.
DR    Pfam; PF01364; Peptidase_C25; 1.
DR    Pfam; PF03785; Peptidase_C25_C; 1.
DR    PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ    SEQUENCE   1723 AA;  186831 MW;  4508A7E50197CEBD CRC64;

Query Match             84.4%;  Score 2094;  DB 2;  Length 1723;
Best Local Similarity   86.2%;  Pred. No. 6.8e-128;
```

Matches 394; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

```
Qy    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
        |||||||||||||||||||||||||||||||||||||      ||  || :||
Db  969 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 1025

Qy   61 SSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
         :  : |||||||| |||||||| :|||||| ||||||||||||||||||||||||||||
Db 1026 YSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 1085

Qy  120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN 179
        ||:|||||| |:||| ||||||| :|.|||||.|||| |||||||||||||.||||.||
Db 1086 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 1143

Qy  180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGGWLCLSSGQLGWLTAHGGT 239
        || ||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db 1144 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGGWLCLSSGQLDWLTAHGGT 1203

Qy  240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFT 299
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1204 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFT 1263

Qy  300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 359
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1264 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 1323

Qy  360 LDDIQFTMGGSPTPDTYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 419
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1324 LDDIQFTMGGSPTPDTYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 1383

Qy  420 PKKCVNVTINPTQFNPVQKNLTGSAVGQKVTLKWDAPN 456
        |||||||| ||||||| ::|   ||:  ||||||:||:
Db 1384 PKKCVNVTINPTQFNPVKNLKAQPDGDGDVVLKWEAPS 1420
```

RESULT 2
Q9R9B7
ID   Q9R9B7
AC   Q9R9B7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin/protease.
GN   HAGE.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Dong H., Progulske-Fox A.;
RT   "Cloning and characterization of hagE from P. gingivalis 381.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF026946; AAD01810.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE 1687 AA; 183702 MW; D085B516A399FE70 CRC64;

Query Match 84.0%; Score 2082; DB 2; Length 1687;
Best Local Similarity 86.2%; Pred. No. 4e-127;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

```
Qy    3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
        ||||||||||||||||||||||||||||||||||||      ||  || :||
Db  935 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 991

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          : |||||||| |||||||| :|||||| |||||||||||||||||||||||||||||||
Db  992 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
        :|||||| |:||| ||||||| :|.|||||.|||| |||||||||||||.||||.|||||
Db 1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy  182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGGWLCLSSGQLGWLTAHGGTNV 241
        :||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 1110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGGWLCLSSGQLDWLTAHGGTNV 1169

Qy  242 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1170 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1229

Qy  302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
        |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db 1230 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1289

Qy  362 DIQFTMGGSPTPDTYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1290 DIQFTMGGSPTPDTYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1349

Qy  422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
        ||||||:.|.||||| ::|   ||:  ||||||:|:
Db 1350 ECVNVTINPTQFNPVKNLKAQPDGDGDVVLKWEAPS 1384
```

RESULT 3
Q51816
ID   Q51816   PRELIMINARY;   PRT;   1704 AA.
AC   Q51816;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arg-gingipain-1 proteinase.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95138080; PubMed=7836351;
RA   Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
RA   Travis J., Barr P.J.;
RT   "Molecular cloning and structural characterization of the Arg-
RT   gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a
RT   proteinase-adhesin polyprotein.";
RL   J. Biol. Chem. 270:1007-1010(1995).
DR   EMBL; U15282; AAA69539.1; -.
DR   PIR; A55426; A55426.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.

DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25_I.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   CHAIN   228   719   MATURE 50-KDA CYSTEINE PROTEINASE
FT   GINGIPAIN.
SQ   SEQUENCE   1704 AA; 185436 MW; 6A34B40131C2A676 CRC64;

Query Match   84.0%; Score 2082; DB 2; Length 1704;
Best Local Similarity   86.2%; Pred. No. 4e-127;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

Qy   3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
            |||||||||||||||||||||||||||||||||||||| ::| ::||||||||
Db   952 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
            ||: ||||| ||||||||||||||||||||||
Db   1009, ESFGLGIGVLTPDNYLIITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVDLPAGTKYVAFRHFGCTDFFWINLD 181
            |||||||||| |||||||||||||||||||||||||||||||||| :|:|||
Db   1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
            ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPTDYYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307 DIQFTMGGSPTPTDYYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy   422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
            ||||| :||: ||||||| ||: |:::
Db   1367 ECVNVTINPTQFNPVKNLKAQPDGQDVVLKWEAPS 1401

RESULT 4
P72194
ID   P72194   PRELIMINARY;   PRT;   1723 AA.
AC   P72194;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97044756; PubMed=8889827;
RA   Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
RT   "Cloning and sequencing of the gene encoding a novel lysine-specific
RT   cysteine proteinase (Lys-gingipain) in Porphyromonas gingivalis:
RT   structural relationship with the arginine-specific cysteine proteinase
RT   (Arg-gingipain).";
RL   J. Biochem. 120:398-406(1996).
EMBL; D83258; BAA11870.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.

DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; DNA ligase.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25_I.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1723 AA; 187261 MW; 5628963D251493EB CRC64;

Query Match   83.8%; Score 2079; DB 2; Length 1723;
Best Local Similarity   85.8%; Pred. No. 6.4e-127;
Matches 392; Conservative 17; Mismatches 42; Indels 6; Gaps 4;

Qy   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
            |||||||||||||||||||||||||||||||||||||||| ::|
Db   969 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 1025

Qy   61 SSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
            : ||: ||||| |||||||||||||||||||||||||
Db   1026 YSESSFGLGIGVLTPDNYLIITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDAS 1085

Qy   120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN 179
            || :||||||| ||||||||||||||||||||||||||||||||:|:|||
Db   1086 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 1143

Qy   180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGT 239
            |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   1144 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGT 1203

Qy   240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 299
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 1263

Qy   300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 359
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1264 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 1323

Qy   360 LDDIQFTMGGSPTPTDYYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 419
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 LDDIQFTMGGSPTPTDYYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 1383

Qy   420 PKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
            ||||||| :||: ||||||| ||: |:::
Db   1384 PKVCVNVTINPTQFNPVKNLKAQPDGQDVVLKWEAPS 1420

RESULT 5
P96967
ID   P96967   PRELIMINARY;   PRT;   1358 AA.
AC   P96967;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin.
GN   HAGD.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Lepine G., Whitlock J., Wojciechowski L., Progulske-Fox A.;
RT   "Cloning, sequencing and characterization of hagD, a member of the
RT   hArep mnltigene family in Porphyromonas gingivalis.";
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
EMBL; U68468; AAB49691.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.

DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0005508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   SMART; SM00060; FN3; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1358 AA; 147102 MW; 47FCA0B25B06DED8 CRC64;

Query Match   83.5%; Score 2071; DB 2; Length 1358;
Best Local Similarity 85.6%; Pred. No. 1.6e-126;
Matches 391; Conservative 17; Mismatches 43; Indels 6; Gaps 4;

Qy   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICA 60
           |||||||||||||||||||||||||||||||||||||||||| :|||
Db   604 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 660

Qy   61 SSASY-INFEGPQNPDNYLVTPELSLPNGGLTFWVCAQDANYASEHYAVYASSTGNDAS 119
           | :. ||| :|| : | :|| |||| |||||||||||||||||||||||||||||
Db   661 YSESFGLGGIGVLTPDNYLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 720

Qy   120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN 179
           || |||| : | .:| : ||||| |||:|:|: ||||||||||||||:|
Db   721 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 778

Qy   180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGCQGWLCLSSGQLGWLTAHGGT 239
           |.: |||||||||||||||||||||||||||||||||||:|||||||||||:|||||||
Db   779 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWDLCLSSGQLDWLTAHGGT 838

Qy   240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 299
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   839 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 898

Qy   300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 359
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   899 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 958

Qy   360 LDDIQFTMGGSPTPDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   959 LDDIQFTMGGSPTPDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 1018

Qy   420 PKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
           ||:.||||:||::|||||||| | :|:|||||||
Db 1019 PKVCVNVTINPTQPNPVKNLKAQPDGGDVVLKWEAPS 1055

RESULT 6
Q51839

RL   Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96311339; PubMed=8713096;
RA   Slakeski N., Cleal S.M., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtR that encodes
RT   an arginine-specific thiol proteinase and multiple adhesins.";
RL   Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Reynolds E.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Slakeski N.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L26341; AAC18876.1; -.
DR   HSSP; P95493; 1CVR.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007107; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease; Signal.
FT   SIGNAL   1   23   POTENTIAL.
FT   CHAIN   228   1706   ARGININE-SPECIFIC THIOL PROTEASE.
SQ   SEQUENCE   1706 AA; 185626 MW; E8BDF07C9813B844 CRC64;

Query Match   83.5%; Score 2070; DB 2; Length 1706;
Best Local Similarity 85.7%; Pred. No. 2.5e-126;
Matches 390; Conservative 18; Mismatches 41; Indels 6; Gaps 4;

Qy   3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASS 62
           ||||||||||||||||||||||||||||||||||||||||||||||||| :|||
Db   954 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
           |: .||| :|| : | :|| |||| ||||||||||||||||||||||||||||
Db 1011 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
           |||||| : | .:| : ||||| |||:|:|: ||||||||||||||:|
Db 1071 TNALLEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGGWLCLSSGQLGWLTAHGGTNV 241
           ||||||||||||||||||||||||||||||||||||:|||||||||||:|||||||||
Db 1129 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWDLCLSSGQLDWLTAHGGTNV 1188

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1249 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy   362 DIQFTMGGSPTPDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1309 DIQFTMGGSPTPDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy   422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456

```
Db  1369 KCVNVTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403
         :|||||::  |||||||::  ||||||:|
```

```
RESULT 7
Q51838
ID   Q51838;      PRELIMINARY;   PRT;   1706 AA.
AC   Q51838;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Protease precursor.
GN   PRPR1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96071894; PubMed=7591131;
RA   Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.;
RT   "Characterization, genetic analysis, and expression of a protease
RT   antigen (PrpRI) of Porphyromonas gingivalis W50.";
RL   Infect. Immun. 63:4744-4754(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.;
RT   "The prpR1 and the prR2 arginine-specific protease genes of
RT   Porphyromonas gingivalis W50 produce five biochemically distinct
RT   enzymes.";
RL   Mol. Microbiol. 23:0-0(1997).
DR   EMBL; X82680; CAA57997.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase C25.
DR   InterPro; IPR005536; Peptidase C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Signal.
FT   SIGNAL      1     23    POTENTIAL.
FT   CHAIN     228    719    ALPHA-PROTEASE.
FT   CHAIN     720   1262    BETA-ADHESIN.
SQ   SEQUENCE  1706 AA;  185705 MW;  0E56DCD87FDA8CDD CRC64;

Query Match
Best Local Similarity  83.0%;  Score 2059;  DB 2;  Length 1706;
Matches  389;  Conservative  17;  Mismatches  43;  Indels  6;  Gaps  4;
                                 Pred. No. 1.3e-125;

Qy    3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS   62
        ||||||||||||||||||||||||||||||||||::  |||||:  ||::
Db  954 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVVS 1010

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDASNF 121
         :|: ||||| |||:||:||:|:|||||:|||||||||||||||||||||||||||||||
Db 1011 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNF 1070

Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
        :|||||:::||| |:||||||| |:|:||:|:||||:||||||||||||::|||:|:||||
Db 1071 TNALLEETITAKG-VRSPEAIRG-RIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy  182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
```

```
Db  1129 EVEIKANGKRADFTETFESSTHGERATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1188
          :||||||||||||||||||||||:  ||||||||||||||||||||||:|||||||||||
Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVANDGFPGDHYAVMISKTGTNAGDFTVV  301
          |:|||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db  1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  1248
Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1249 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308
Qy   362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1309 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368
Qy   422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  456
          :|||||:  |||||:||:  :  ::||.:|::|
Db  1369 KCVNVTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403
```

```
RESULT 8
Q92NB5
ID   Q92NB5;      PRELIMINARY;   PRT;   1223 AA.
AC   Q92NB5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   130k-HMGD (Fragment).
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381.
RX   MEDLINE=99143166; PubMed=9988746;
RA   Shibata Y., Hayakawa M., Takiguchi H., Shiroza T., Abiko Y.;
RT   "Determination and characterization of the hemagglutinin-associated
RT   short motifs found in Porphyromonas gingivalis multiple gene
RT   products.";
RL   J. Biol. Chem. 274:5012-5020(1999).
DR   EMBL; AB019363; BAA34341.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   NON_TER     1      1
SQ   SEQUENCE  1223 AA;  131542 MW;  0025CD2BA9F91B3 CRC64;

Query Match
Best Local Similarity  82.8%;  Score 2054.5;  DB 2;  Length 1223;
                               Pred. No. 1.6e-125;
Matches  391;  Conservative  15;  Mismatches  44;  Indels  9;  Gaps  6;

Qy    1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC  59
        |||||||||||| |:||||||||||||||||||||||::  |||||:  |::
Db  460 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC  516

Qy   60 ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDA 118
          :|: :|: ||||| |||:||:||:|:|||||:||||||||||||||||||||||||||
Db  517 VYSESFGLGGIGVLTPDNVLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDA 576

Qy  119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
        |||:|||||:::||| |:||||||| |:|:||:|:||||:||||||||||||::|||:|:|
Db  577 SNFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 634
```

```
Qy   391 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 450
             ||||||||||||||||||||||||||||||||||||||||||
Db   119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
             : : ||| :||| :||  : ||| ::::|| :|||:||||||||||||||| :.:| :|
Db   451 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 508
Qy   179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG 238
            :|| |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   509 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGG 568
Qy   239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
            :|||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   569 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 628
Qy   299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
            ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db   629 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 688
Qy   359 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   689 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 748
Qy   419 SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 455
            |||:||||||.: :|||||||||:  : :    |.||:|.:.
Db   749 SPKKCVNVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 787
```

RESULT 10
```
O07442
ID   O07442             PRELIMINARY;        PRT;  1732 AA.
AC   O07442;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine-specific cysteine proteinase.
GN   PRTK.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=99235907; PubMed=10219167;
RA   Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtK that encodes
RT   a lysine-specific cysteine proteinase and three sequence-related
RT   adhesins.";
RL   Oral Microbiol. Immunol. 14:92-97(1999).
DR   EMBL; U75366; AAB60809.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE 1732 AA; 187914 MW; 45D5B91377391703 CRC64;

Query Match             82.2%;  Score 2039.5;  DB 2;  Length 1732;
Best Local Similarity   84.5%;  Pred. No. 2.4e-124;
Matches  388;  Conservative   18;  Mismatches   44;  Indels    9;  Gaps    6;

Qy     1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
           ||||||||||||| |||||||||||||||||||||||||           |||.|||.|
Db   969 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025
```

```
Qy   179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG 238
            :|| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   635 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 694
Qy   239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   695 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 754
Qy   299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
            ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db   755 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 814
Qy   359 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   815 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 874
Qy   419 SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 455
            |||:||||||.: :|||||||||:  :.:    |.||:|.:.
Db   875 SPKVCVNVTINPTQFNPVQNLTAEQAPNSMDAILKWNAP 913
```

RESULT 9
```
P72196
ID   P72196             PRELIMINARY;        PRT;  1097 AA.
AC   P72196;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   TonB-linked adhesin precursor.
GN   TLA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97386416; PubMed=9244265;
RA   Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA   Curtis M.A.;
RT   "The tla gene of Porphyromonas gingivalis W50: a homologue of the
RT   arginine-specific protease precursor (PrpRI) which shares sequence
RT   similarity to TonB-linked receptors.";
RL   J. Bacteriol. 179:4778-4788(1997).
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR   EMBL; Y07618; CAA68897.1; -.
DR   GO; GO:0019867; C:outer membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Membrane; Outer membrane; Receptor; Signal; TonB box.
FT   SIGNAL        1     53       POTENTIAL.
SQ   SEQUENCE 1097 AA; 118731 MW; 73BBA337B421F8B9 CRC64;

Query Match             82.2%;  Score 2039.5;  DB 2;  Length 1097;
Best Local Similarity   84.5%;  Pred. No. 1.3e-124;
Matches  388;  Conservative   18;  Mismatches   44;  Indels    9;  Gaps    6;

Qy     1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
           ||||||||||||| |||||||||||||||||||||||||           |||.|||.|
Db   334 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 390
Qy    60 ASSASY-INFEGQPQNPDNYLVTPELSLPNGGTLFWVCAQDANYASEHYAVYASSTGNDA 118
           :  : :
```

```
Qy    60 ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
            ||:::  ||||| |:  |||||||: ::|||||||||||||||||||||||||||||||
Db  1026 VYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy   119 SNFANALLEEVLTAKTVVTAEPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
           |||:||||||:: ||| ||: ||||| | |||| |||||:||||||||||||||:|||:|
Db  1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy   179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG 238
           :||:|||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db  1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy   239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
           :|||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy   299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
           |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db  1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy   359 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1324 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy   419 SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 455
           |||:||:|||   :||||||||:::      |:||||:||
Db  1384 SPKKCVNVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422
```

RESULT 11
```
ID   Q51817                       PRELIMINARY;      PRT;   1732 AA.
AC   Q51817;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Porphypain.
GN   PRTP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W12;
RX   MEDLINE=96213011; PubMed=8631659;
RA   Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA   Progulske-Fox A., Lantz M.S.;
RT   "Analysis of the prtP gene encoding porphypain, a cysteine proteinase
RT   of Porphyromonas gingivalis.";
RL   J. Bacteriol. 178:2734-2741(1996).
DR   EMBL; U42210; AAB06565.1; -.
DR   PIR; T30836; T30836.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE 1732 AA; 187875 MW; 654271DBEF7BCAE4 CRC64;

Query Match             82.0%;  Score 2034.5;  DB 2;  Length 1732;
  Best Local Similarity   84.3%;
  Matches  387;  Conservative   19;  Mismatches   44;  Indels    9;  Gaps    6;
```

Right column:

```
Qy     1 GTPNPNPNPNGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAIC 59
           |||||||||| |  ||||||||||||||||||||||||   |  |||||  :|  |:|:|
Db   969 GTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy    60 ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
           ||:::  ||||| |:  |||||||: ::|||||||||||||||||||||||||||||||
Db  1026 VYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy   119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
           |||:||||||:: ||| ||: ||||| | |||| |||||:||||||||||||||:|||:|
Db  1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy   179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG 238
           :||:|||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db  1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy   239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
           :|||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy   299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
           |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db  1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy   359 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1324 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy   419 SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 455
           |||:||:|||   :||||||||:::      |:||||:||
Db  1384 SPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422
```

RESULT 12
```
ID   O52050                       PRELIMINARY;      PRT;   1732 AA.
AC   O52050;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine specific cysteine protease.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=98298016; PubMed=9632563;
RA   Lewis J.P., Macrina F.L.;
RT   "IS195, an insertion sequence-like element associated with protease
RT   genes in Porphyromonas gingivalis.";
RL   Infect. Immun. 66:3035-3042(1998).
DR   EMBL; AF017059; AAC26523.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE 1732 AA; 187931 MW; B2337463D5CB5EA5 CRC64;

Query Match             81.5%;  Score 2020.5;  DB 2;  Length 1732;
                                                              6;
```

Best Local Similarity   83.9%;   Pred. No. 4.2e-123;
Matches 385;   Conservative 19;   Mismatches 46;   Indels 9;   Gaps 6;

```
Qy      1 GTPNPNPNPDGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
            |||||||||||| :||||||||||||||||||||||||||||||||||||:||||||
Db    969 GTPNPNPNPNPGTTISESFENGIPASWKTIDADGDGHGHWKPGNAPG---IAGYNSNGC 1025

Qy     60 ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
            ||||| ::||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1026 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy    119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
            |||||||||||| |: |||| |||||: ||||||| ::|||||||||||||||| |:|
Db   1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy    179 NLDDVEIKANGKRADFTETFSSSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG 238
            :||||||||||||||||||||||:|||||||||||||||||||||||||||||:|||||
Db   1144 DLDEVEIKANGKRADFTETFSSSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy    239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
            :||| ||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy    299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
            |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy    359 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
            ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   1324 LLDDIQFTMGGSPTPTDYTYTVYRDGTNIKEGLTETTFEEYGVATGNHEYCVEVKYTAGV 1383

Qy    419 SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 455
            ||:|| |:|||| ||||||||::| |  |||:|||:||:
Db   1384 SPKKCVDVTVNSTQPNPVQNLTAEQAPNSMDAILKWNAP 1422
```

RESULT 13
Q9F4J0
```
ID   Q9F4J0;                 PRELIMINARY;      PRT;      925 AA.
AC   Q9F4J0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative outer membrane protein PG57.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA   Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT   "Identification of vaccine candidates from genomic analysis of
RT   Porphyromonas gingivalis.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY007522; AAG24228.1; -.
DR   InterPro; IPR003961; FN_III.
DR   SMART; SM00060; FN3; 1.
SQ   SEQUENCE  925 AA;  103632 MW;  5FF2198D6914DAE8 CRC64;
```

Query Match   27.5%;   Score 682;   DB 2;   Length 925;
Best Local Similarity   30.2%;   Pred. No. 4e-36;
Matches 176;   Conservative 62;   Mismatches 132;   Indels 212;   Gaps 22;

```
Qy     16 LSESFENG-IPASWKTIDADGDGDGNNWTTTPPPGG-----TSFAGHNSAICASSASYINF 68
```

```
Db    311 |:|:||||:|:||||||||||| ||||||:|----------GHYLNAYDAFPGHNGGHCSLSASYVPG 363

Qy     69 EGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEE 128
            :|||||:||:||| :|||:||||:|||::| ||:|| ||||:| |
Db    364 IGPVTPDNYLITPKVE--GAKRVKYWVSTQDANWAAEHYAVMASTTGTAVGDFV-ILFEE 420

Qy    129 VLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV----- 183
            : :|||||||| |:|||::|||| | |:||||||||||||:|| || :::|||
Db    421 TMTAKPT-----------------GAWYERTINLPEGTKYIAWRHYNCTDIYFLKLDDITVFGT 467

Qy    184 ---------------------------------------------EIKANG--------- 189
                                                          :|:||||
Db    468 PASEPEVTDFVVSLIENNKGRLKWNYPNGYEPDKTDKDPLQLAGYNIYANGSLLVHIQ 527

Qy    190 -----------KRADFTET------------------------------------FESSTHGE 205
                       |||||||
Db    528 DPTVLEYIDETYSSRDDQVEVEYCVTAVYNDNIESQSVCDKLIYDSQSDIIIYEGFEAGS 587

Qy    206 APAEWTTIDADGDGQGWLCLSSGQLGW-LTAHGGTNVVASFSWNGM-ALNPDNYLISKD 262
            |||:||||||||||| |:||||:|||  ||||||||||||||||| ||||||||::|:
Db    588 IPEGWLLIDADGDCNVNW------DYYPWTMYGHDSEKCIASPSYLPMIGVLTPDNYLVTPR 642

Qy    263 VTGATKVKYYYAVNDG-FPGDHYAVMISKIGTNAGDFTVVFEETPNGINKGGARFGLSTE 321
            :||||||||:||||| :::|| |||||| :|
Db    643 LEGAKLVKYWVSAQDAVYSAEHYAVMVSTTGTAVEDFVLLFEET------MTAK 690

Qy    322 ADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDI--------------YTYTVYRDGT-- 363
            |:|||: :||||||||:||||||||||||||||||||||||
Db    691 ANGA------WYERTITLPAGTKYIAWRHYDCTDMFLLLDDITVYRSTETVPEPVTDFVV 745

Qy    364 ------------------------QFTMGGSPTPTD------YTYTVYRDGT--KIKEGLTETTF 396
                                    ::||:|||||||        ::|:||
Db    746 SLIENNKGRLKWNYPNGYEPDKTDKKPLQLTGYNIYANGSLLVHIQDPTVLEYIDETYS 805

Qy    397 EEDGVATGNHEYCVEVKYTAGV-SPKEC--VNVTV--DPVQ 432
            ::||||||||||||||||||||:::|||:
Db    806 SRDGGQV--EMEYCVTAVYNDNIESQSVCDKLNYTITSLDNIQ 845
```

RESULT 14
Q9KIB3
```
ID   Q9KIB3;                 PRELIMINARY;      PRT;      312 AA.
AC   Q9KIB3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical outer membrane protein PG27.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C.;
RT   "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF237555; AAF81413.1; -.
SQ   SEQUENCE  312 AA;  34592 MW;  0D5792C9643A25F5 CRC64;
```

Query Match   22.6%;   Score 561.5;   DB 2;   Length 312;
Best Local Similarity   32.8%;   Pred. No. 7e-29;
Matches 136;   Conservative 73;   Mismatches 23;   Indels 183;   Gaps 8;

```
Qy     16 LSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYINFEGPQNPD 75
```

```
Db      4 LSESFESGIPAIWKTIDADGDYNWMHL-----TNFTG-QSGLCVSSASYIGGVGALTPD  57

Qy     76 NYLVTPELSLPNGG--TLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAK 133

Db     58 NYLITPELKLPTDALVEIIYWVCTQDLTAPSEHYAVYSSSTGNNAADFVNLIYEETLTAK 117

Qy    134 TVVTAPEAIRGTRVOGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGKRAD 193

Db    118 R-IQSPELIRGNRTQGVWYQRKVVLPNDTKYVAFRHFNSTDNFWLNLDEVSIL------ 169

Qy    194 FTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWNGMALN 253

Db    170 ----------------------------------------------------------- 169

Qy    254 PDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGG 313

Db    170 ----------------------------------------------------YTP----- 172

Qy    314 ARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTP 373

Db    173 ------------------------LP------RRAPC------------------PHP 182

Qy    374 TDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTV 428

Db    183 GGYTYSVFRDGQKIASGLSALAYIDTDVPYGTQDYCVQVNYLQGDSYKVCKNIVV 237
```

RESULT 15
```
Q9XBU9
ID   Q9XBU9          PRELIMINARY;        PRT;     293 AA.
AC   Q9XBU9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Immunoreactive 32 kDa antigen PG25.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF155351, AAD38410.1; -.
SQ   SEQUENCE   293 AA;   32272 MW;   CC03EAC241F7F6F1 CRC64;
```

```
Query Match              9.4%;   Score 233;  DB 2;  Length 293;
Best Local Similarity   59.0%;   Pred. No. 1.7e-07;
Matches 46;   Conservative 7;   Mismatches  23;   Indels   2;   Gaps   1;

Qy    376 YTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNP 435
              |||||||||  |::||  |  |||||||||::::|  ||||||||::||  |  ||||
Db     77 YTYTYVRDGVKIASGLTEPTFLDEDVPAGEHTYCVEVQYQGGVSDKVCVDVEVK--DFKP 134

Qy    436 VQNLTGSAVGQKVTLKWD 453
              :|||  |  : ||
Db    135 VTNLTGTASNDEVSLDWD 152
```

Search completed: May 18, 2004, 11:46:38
Job time : 35.3368 secs

OM protein - protein search, using sw model

          GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: May 18, 2004, 11:38:40 ; Search time 14.0888 Seconds
                              (without alignments)
                              1670.936 Million cell updates/sec

Title: US-08-570-311-18
Perfect score: 2480
Sequence: 1 GTPNPNPNPNPGTTTLSESF........QNLTGSAVGQKVTLKWDAPN 456

Scoring table: BLOSUM62
          Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 5162597I residues

Total number of hits satisfying chosen parameters: . 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued Patents AA:*
         1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
         2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
         3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
         4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
         5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
         6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                           SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	2480	100.0	2480	2	US-08-570-311-18	Sequence 18, Appl
2	2480	100.0	456	2	US-08-570-311-20	Sequence 20, Appl
3	2480	100.0	2628	2	US-08-570-311-14	Sequence 14, Appl
4	2436	98.2	450	2	US-08-570-311-16	Sequence 16, Appl
5	2351	94.8	439	2	US-08-570-311-22	Sequence 22, Appl
6	2082	84.0	1477	4	US-09-482-500A-1	Sequence 1, Appl
7	2082	84.0	1687	2	US-08-570-311-29	Sequence 29, Appl
8	2082	84.0	1704	4	US-08-336-308A-10	Sequence 6, Appl
9	2082	84.0	1704	3	US-08-822-324-6	Sequence 6, Appl
10	2082	84.0	1704	3	US-09-490-931-10	Sequence 10, Appl
11	2070	83.5	1706	3	US-09-066-330-10	Sequence 10, Appl
12	2066	83.3	1087	2	US-08-570-311-8	Sequence 8, Appl
13	2066	83.3	1358	2	US-08-353-485-8	Sequence 8, Appl
14	2066	83.3	1732	2	US-08-570-311-27	Sequence 27, Appl
15	2039.5	82.2	1732	4	US-09-066-330-11	Sequence 11, Appl
16	2034.5	82.0	1732	2	US-08-570-311-10	Sequence 10, Appl
17	2034.5	82.0	497	2	US-08-353-485-10	Sequence 10, Appl
18	546.5	22.0	497	2	US-08-570-311-2	Sequence 2, Appl
19	546.5	22.0	942	2	US-08-353-485-2	Sequence 2, Appl
20	296.5	12.0	942	1	US-08-141-324-14	Sequence 14, Appl
21	296.5	12.0	942	1	US-08-541-902-14	Sequence 14, Appl
22	236	9.5	49	3	US-08-822-324-18	Sequence 18, Appl
23	189	7.6	46	3	US-08-822-324-9	Sequence 9, Appl
24	137	5.5	1833	4	US-08-621-944A-4	Sequence 4, Appl
25	137	5.5	1833	4	US-08-945-567D-4	Sequence 4, Appl
26	137	5.5	1992	4	US-08-621-944A-3	Sequence 3, Appl
27	137	5.5	1992	4	US-08-945-567D-3	Sequence 3, Appl
28	134.5	5.4	2736	4	US-09-252-991A-30227	Sequence 30227, A
29	132.5	5.3	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
30	131.5	5.3	492	4	US-09-482-500A-2	Sequence 2, Appli
31	131.5	5.3	737	1	US-08-119-361-5	Sequence 5, Appli
32	131.5	5.3	737	3	US-08-336-308A-4	Sequence 4, Appli
33	131.5	5.3	737	3	US-08-822-324-4	Sequence 4, Appli
34	131.5	5.3	737	3	US-09-490-931-4	Sequence 4, Appli
35	131.5	5.3	2123	3	US-08-968-685A-10	Sequence 10, Appl
36	129.5	5.2	1638	4	US-09-071-035-258	Sequence 258, App
37	129.5	5.2	1638	4	US-09-071-035-262	Sequence 262, App
38	129.5	5.2	1638	4	US-09-071-035-266	Sequence 266, App
39	129.5	5.2	1747	4	US-09-134-000C-5999	Sequence 5999, Ap
40	129	5.2	24	3	US-09-066-330-6	Sequence 6, Appli
41	128	5.2	25	3	US-08-822-324-19	Sequence 19, Appl
42	126	5.1	2411	4	US-09-268-347-36	Sequence 36, Appl
43	123.5	5.0	669	4	US-09-071-035-264	Sequence 264, App
44	123.5	5.0	2048	4	US-09-268-347-48	Sequence 48, Appl
45	121	4.9	509	3	US-08-822-324-8	Sequence 8, Appli

                           ALIGNMENTS

RESULT 1
US-08-570-311-18
; Sequence 18, Application US/0857031I
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tunwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:

```
;   LENGTH: 456 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: Protein
US-08-570-311-18

Query Match       100.0%;   Score 2480; DB 2; Length 456;
Best Local Similarity   100.0%;   Pred. No. 2.8e-204;
Matches 456;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy    1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA  60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361 DDIQFTMGGSPTPTDYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

RESULT 2
US-08-570-311-20
; Sequence 20, Application US/08570311
; Patent No. 5824791
GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: .424

;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 456 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-20

Query Match   100.0%;   Score 2480;   DB 2;   Length 456;
Best Local Similarity   100.0%;   Pred. No. 2.8e-204;
Matches 456;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy    1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA  60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361 DDIQFTMGGSPTPTDYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

RESULT 3
US-08-570-311-14
; Sequence 14, Application US/08570311
; Patent No. 5824791
GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn

```
Qy   361 DDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1310 DDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1369

Qy   421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
         ||||||||||||||||||||||||||||||||||||
Db  1370 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 1405
```

RESULT 4
US-08-570-311-16
; Sequence 16, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 16:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 450 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-16

Query Match   98.2%; Score 2436; DB 2; Length 450;
Best Local Similarity   99.8%; Pred. No. 1.6e-200;
Matches 449; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   7 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 66
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60
```

APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ted W. Whitlock
 STREET: 2421 N.W. 41st Street, Suite A-1
 CITY: Gainesville
 STATE: FL
 COUNTRY: USA
 ZIP: 32606
COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/570,311
 FILING DATE:
 CLASSIFICATION: 424
PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/353,485
 FILING DATE: 09-DEC-1994
 CLASSIFICATION: 424
PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/647,119
 FILING DATE: 25-JAN-1991
 CLASSIFICATION: 424
PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/241,640
 FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
 NAME: Whitlock, Ted W.
 REGISTRATION NUMBER: 36,965
 REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
 TELEPHONE: (904) 375-8100
 TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
  LENGTH: 2628 amino acids
  TYPE: amino acid
  TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-570-311-14

```
Query Match             100.0%;  Score 2480;  DB 2;  Length 2628;
Best Local Similarity   100.0%;  Pred. No. 3.8e-203;
Matches  456;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  950 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 1009

Qy   61 SSASYINFEGPQNPDNYLIVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1010 SSASYINFEGPQNPDNYLIVTPELSLPNGGTLTFWVCAQDANYASEHYAVIASSTGNDASN 1069

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAPRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1070 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAPRHFGCTDFFWINL 1129

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1130 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 1189

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1190 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 1249

Qy  301 VFEETPNGIINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1250 VFEETPNGIINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1309
```

REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (904) 375-8100
    TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 22:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 439 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-570-311-22

Query Match   94.8%; Score 2351; DB 2; Length 439;
Best Local Similarity 98.2%; Pred. No. 3e-193;
Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy   1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGGNNWTTTPPPGGTSFAGHNSAICA 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGGNNWTTTPPPGGTSFAGHNSAICV 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy 121 FANALLEEVLTAKTVVTAPEAIRGTRVGQTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 FANALLEEVLTAKTVVTAPEAIRGTRVGQTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy 181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
         |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240

Qy 241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy 301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy 361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy 421 KECVNVTVDPVQFNPVQNL 439
         | :||||||||||||||||
Db 421 KVCVNVTINPTQFNPVQNL 439

Query Match   84.0%; Score 2082; DB 4; Length 1477;

Qy   67 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 NFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120

Qy 127 EEVLTAKTVVTAPEAIRGTRVGQTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 186
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 EEVLTAKTVVTAPEAIRGTRVGQTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180

Qy 187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFS 246
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240

Qy 247 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300

Qy 307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360

Qy 367 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 426
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420

Qy 427 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
         ||||||||||||||||||||||||||||||
Db 421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450

Best Local Similarity 86.2%; Pred. No. 2.1e-169; Indels 6; Gaps 4;
Matches 392; Conservative 17; Mismatches 40;

```
Qy    3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
         |||||||||||||||||||||||||||||||||| :|: ||  ||
Db  725 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 781

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
         :| ||||||||||||||||||||||||||||||||||||||||
Db  782 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 841

Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
         ||||| ||| :||   ||||   |:||:|| ||||||||||||||||  :|:|
Db  842 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 899

Qy  182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  900 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 959

Qy  242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  960 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1019

Qy  302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1020 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1079

Qy  362 DIQFTMGGSPTPTDYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1080 DIQFTMGGSPTPTDYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1139

Qy  422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
         |||||||| :||   ||||
Db 1140 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1174
```

RESULT 7
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-29

Query Match 84.0%; Score 2082; DB 2; Length 1687;
Best Local Similarity 86.2%; Pred. No. 2.5e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

```
Qy    3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
Db  935 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 991

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db  992 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
Db 1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy  182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
Db 1110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1169

Qy  242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
Db 1170 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1229

Qy  302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
Db 1230 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1289

Qy  362 DIQFTMGGSPTPTDYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
Db 1290 DIQFTMGGSPTPTDYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1349

Qy  422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
Db 1350 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1384
```

RESULT 8
US-08-336-308A-10
; Sequence 10, Application US/0833630BA
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado

GENERAL INFORMATION:
;   APPLICANT: Potempa, Jan S.
;   APPLICANT: Travis, James
;   APPLICANT: Genco, Caroline A.
;   TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
;   TITLE OF INVENTION: METHODS
;   NUMBER OF SEQUENCES: 24
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
;   STREET: 5370 Manhattan Circle, Suite 201
;   CITY: Boulder
;   STATE: CO
;   COUNTRY: US
;   ZIP: 80303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/822,324
;   FILING DATE: 21-MAR-1997
;   CLASSIFICATION: 536
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/013,945
;   FILING DATE: 22-MAR-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ferber, Donna M.
;   REGISTRATION NUMBER: 33,878
;   REFERENCE/DOCKET NUMBER: 103-95 WO
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (303) 488-8080
;   TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1704 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-822-324-6

Query Match   84.0%; Score 2082; DB 3; Length 1704;
Best Local Similarity 86.2%; Pred. No. 2.6e-169;
Matches   392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

Qy   3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASS 62
          ||||||||||||||||||||||||||||||||||: |:||
Db   952 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          :|: | |:|||||:|| | ||||| |||||||||||||||||||||||||||||||
Db   1009 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTMYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          :||||| |||| | :|:||| |:|||:||||:||||||||||||||| ::|:|:|
Db   1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
          :||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db   1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db   1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307 DIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

RESULT 9
US-08-822-324-6
; Sequence 6, Application US/08822324
; Patent No. 6129917

```
Db  1127  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNV 1186

Db   242  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
Db  1187  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
Db  1247  FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362  DIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 421
Db  1307  DIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy   422  ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
Db  1367  ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

RESULT 11
US-09-066-330-10
; Sequence 10, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275
; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; SOFTWARE: PatentIn Ver. 2.0
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-10

Query Match   83.5%; Score 2070; DB 4; Length 1706;
Best Local Similarity   85.7%; Pred. No. 2.7e-168;
Matches 390; Conservative 18; Mismatches 41; Indels 6; Gaps 4;

```
Qy    3  PNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASS 62
Db  954  PNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy   63  ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 1011  ESFGLGGIGVLTPDNVLITPALDLPNGGKLTFWVCAQDANYASEHYAVIASSTGNDASNF 1070

Qy  122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
Db 1071  TNALLEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy  182  DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTNV 241
Db 1129  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNV 1188

Qy  242  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
Db 1189  VSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy  302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
Db 1249  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy  362  DIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 421
```

```
Qy   422  ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
Db  1367  ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

RESULT 10
US-09-490-931-10
; Sequence 10, Application US/09490931
; Patent No. 6274718
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/490,931
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/336,308
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-490-931-10

Query Match   84.0%; Score 2082; DB 3; Length 1704;
Best Local Similarity   86.2%; Pred. No. 2.6e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

```
Qy    3  PNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASS 62
Db  952  PNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG--IAGYNSNGCVYS 1008

Qy   63  ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 1009  ESFGLGGIGVLTPDNVLITPALDLPNGGKLTFWVCAQDANYASEHYAVIASSTGNDASNF 1068

Qy  122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
Db 1069  TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy  182  DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTNV 241
```

```
Db   390 YSESFGLGGIGVLTPDNYLIITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 449

Qy   120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN 179
         :|:||||  |:  ::||||  :   ||  |::   ||||||| |:|::
Db   450 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 507

Qy   180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGT 239
         ||:||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   508 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGT 567

Qy   240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 299
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   568 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 627

Qy   300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 359
         |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db   628 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYIL 687

Qy   360 LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 419
         ||||||||||||||||||||||||:||||| |:||||||||||||||||||||||||||
Db   688 LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 747

Qy   420 PKECVNTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
         |||||||  :|||||:|| ||||| :::||:|
Db   748 PKVCVNTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 784
```

```
RESULT 13
US-08-353-485-8
; Sequence 8, Application US/08353485
; Patent No. 5830710
; GENERAL INFORMATION:
;  APPLICANT: Progulske-Fox, Ann
;  APPLICANT: Tumwasorn, Somying
;  APPLICANT: Lepine, Guylaine
;  APPLICANT: Han, Naiming
;  APPLICANT: Lantz, Marilyn
;  APPLICANT: Patti, Joseph
;  TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;  TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;  NUMBER OF SEQUENCES: 12
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C2
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
;  SEQUENCE CHARACTERISTICS:
```

```
Db  1309 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy   422 ECVNTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
         :|||||:| |||||:||: ||||: | ||:||:|:
Db  1369 KCVNTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 12
US-08-570-311-8
; Sequence 8, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;  APPLICANT: Progulske-Fox, Ann
;  APPLICANT: Tumwasorn, Somying
;  APPLICANT: Lepine, Guylaine
;  APPLICANT: Han, Naiming
;  APPLICANT: Lantz, Marilyn
;  APPLICANT: Patti, Joseph
;  TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;  TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;  NUMBER OF SEQUENCES: 29
;  CORRESPONDENCE ADDRESS:
;   ADDRESSE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 1087 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;  MOLECULE TYPE: protein
US-08-570-311-8

Query Match             83.3%;  Score 2066;  DB 2;  Length 1087;
Best Local Similarity   85.1%;  Pred. No. 3.1e-168;
Matches  390;  Conservative  18;  Mismatches  43;  Indels  6;  Gaps   4;

Qy     1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNWTTTPPPGGTSFAGHNSAICA  60
         |||||||||||||||||||||||||||||||||||||||  :||:
Db   333 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 389

Qy    61 SSASY-INFEGPQNPDNYLVTPELSLPNGGLTFWVCAQDANYASEHYAVYASSTGNDAS 119
         ||:  : | |  ||||||| ||||||||||||||||||||||||||||||||
```

```
;   LENGTH: 1087 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-353-485-8

Query Match            83.3%;   Score 2066; DB 2; Length 1087;
Best Local Similarity  85.3%;   Pred. No. 3.1e-168;
Matches 390; Conservative 18; Mismatches 43; Indels 6; Gaps 4;

Qy      1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA 60
          |||||||||||||||||||||||||||||||||||||||||:||   |||:||
Db    333 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 389

Qy     61 SSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
          |||:: ||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db    390 YSESFGLGIGVLTPDNYLIITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 449

Qy    120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN 179
          ||||||||||:|||||||||||||||:|:||||||||||||||||||||||:|:
Db    450 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 507

Qy    180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGT 239
          ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    508 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGT 567

Qy    240 NVVASFSWNGMALNPDNYLLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    568 NVVASFSWNGMALNPDNYLLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 627

Qy    300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 359
          ||||||||||||||||||||||:|||||||||||||||||||||||||||||||:|||
Db    628 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYIL 687

Qy    360 LDDIQFTMGGSPTPDTYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    688 LDDIQFTMGGSPTPDTYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 747

Qy    420 PKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||:|||:||||:|||  |||:|:||||:||||||:
Db    748 PKVCVNVTINPTQFNPVKNLKAQPDGDVVLKWEAPS 784
```

```
RESULT 14
US-08-570-311-27
; Sequence 27, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Ted W. Whitlock
;     STREET: 2421 N.W. 41st Street, Suite A-1
;     CITY: Gainesville
;     STATE: FL
;     COUNTRY: USA
;     ZIP: 32606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentin Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/570,311
;     FILING DATE:
;     CLASSIFICATION: 424
```

```
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/353,485
;     FILING DATE: 09-DEC-1994
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/647,119
;     FILING DATE: 25-JAN-1991
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/241,640
;     FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Whitlock, Ted W.
;     REGISTRATION NUMBER: 36,965
;     REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (904) 375-8100
;     TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 27:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1358 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-570-311-27

Query Match            83.3%;   Score 2066; DB 2; Length 1358;
Best Local Similarity  85.3%;   Pred. No. 4.3e-168;
Matches 390; Conservative 18; Mismatches 43; Indels 6; Gaps 4;

Qy      1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA 60
          |||||||||||||||||||||||||||||||||||||||:||   |||:||
Db    604 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 660

Qy     61 SSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
          |||:: ||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 YSESFGLGIGVLTPDNYLIITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 720

Qy    120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN 179
          ||||||||||:|||||||||||||||:|:||||||||||||||||||||||:|:
Db    721 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 778

Qy    180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGT 239
          ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    779 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGT 838

Qy    240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    839 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 898

Qy    300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 359
          ||||||||||||||||||||||:||||||||||||||||||||||||||||||:|||
Db    899 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYIL 958

Qy    360 LDDIQFTMGGSPTPDTYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    959 LDDIQFTMGGSPTPDTYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 1018

Qy    420 PKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||:|||:||||:|||  |||:|:||||:||||||:
Db   1019 PKVCVNVTINPTQFNPVKNLKAQPDGDVVLKWEAPS 1055
```

```
RESULT 15
US-09-066-330-11
; Sequence 11, Application US/09066330A
; Patent No. 651666
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric C.
;   APPLICANT: Bhogal, Peter S.
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;   FILE REFERENCE: Reynolds
```

us-08-570-311-18.rai

```
;    CURRENT APPLICATION NUMBER: US/09/066,330A
;    CURRENT FILING DATE: 1998-09-15
;    EARLIER APPLICATION NUMBER: PN 6275
;    EARLIER FILING DATE: 1995-10-30
;    EARLIER APPLICATION NUMBER: PCT/AU96/00673
;    EARLIER FILING DATE: 1996-10-30
;    NUMBER OF SEQ ID NOS: 15
;    SOFTWARE: PatentIn Ver. 2.0
;    SEQ ID NO 11
;      LENGTH: 1732
;      TYPE: PRT
;      ORGANISM: Porphyromonas gingivalis
US-09-066-330-11
```

```
Query Match        82.2%;      Score 2039.5; DB 4;   Length 1732;
Best Local Similarity    84.5%;    Pred. No. 1.2e-165;
Matches   388;  Conservative   18;  Mismatches   44;  Indels    9;  Gaps     6;

Qy     1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC   59
           ||||||||||| | ||||||||||||||||||||||||         :  || |
Db   969 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy    60 ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA  118
           | |: : |  |||  | | |||||| : |||||||||||||||||||||||||||||||
Db  1026 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy   119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI  178
           ||| |||| | :: ||   : |   ||| |||||||||||||||||||||||||  :  :
Db  1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy   179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG  238
           :||:|||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy   239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF  298
           :||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy   299 TVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYYAFRHYNCSDLNYI  358
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db  1264 TVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy   359 LLDDIQFTMGGSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV  418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1324 LLDDIQFTMGGSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy   419 SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP  455
           |||:|||||| ||||||||||||   |  || |:|||||
Db  1384 SPKKCVNVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422
```

Search completed: May 18, 2004, 11:49:14
Job time : 15.0888 secs

```
              GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 18, 2004, 11:42:50 ; Search time 35.3211 Seconds
                                     (without alignments)
                                     3592.387 Million cell updates/sec

Title:           US-08-570-311-18
Perfect score:   2480
Sequence:        1 GTPNPNPNPGTTLSESF.........QNLTGSAVGQXVTLKWDAPN 456

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters:  ,   1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published Applications AA:*
         1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
         2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
         3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
         4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
         5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
         6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
         7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
         8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
         9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
        10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
        11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
        12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
        13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
        14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
        15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
        16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
        17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
        18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                         SUMMARIES
```

Result No.	Score	% Query Match	DB	Length	ID	Description
1	2070	83.5	14	1706	US-10-229-066-10	Sequence 10, Appl
2	2039.5	82.2	14	1732	US-10-229-066-11	Sequence 11, Appli
3	726	29.3	15	419	US-10-174-695-3	Sequence 3, Appli
4	710.5	28.6	15	419	US-10-174-695-5	Sequence 5, Appli
5	385	15.5	15	231	US-10-174-695-6	Sequence 4, Appli
6	232.5	9.4	15	196	US-10-174-695-4	Sequence 4, Appli
7	153	6.2	15	29	US-10-387-977-18	Sequence 18, Appl
8	153	6.2	15	29	US-10-387-977-21	Sequence 21, Appl
9	153	6.2	15	29	US-10-387-977-23	Sequence 23, Appl
10	153	6.2	15	29	US-10-387-977-24	Sequence 24, Appl
11	153	6.2	15	29	US-10-387-977-26	Sequence 26, Appl
12	147	5.9	15	29	US-10-387-977-17	Sequence 17, Appl
13	147	5.9	12	872	US-10-282-122A-55467	Sequence 55467, A
14	145	5.8	15	29	US-10-387-977-25	Sequence 25, Appl
15	143	5.8	15	509	US-10-387-977-101	Sequence 101, App

16	137	5.5	14	1833	US-10-175-282-4	Sequence 4, Appli
17	137	5.5	14	1833	US-10-175-275-4	Sequence 4, Appli
18	137	5.5	14	1992	US-10-175-282-3	Sequence 3, Appli
19	137	5.5	14	1992	US-10-175-275-3	Sequence 3, Appli
20	136	5.5	15	29	US-10-387-977-20	Sequence 20, Appl
21	134	5.4	15	25	US-10-387-977-77	Sequence 77, Appl
22	134	5.4	15	25	US-10-387-977-80	Sequence 80, Appl
23	134	5.4	15	27	US-10-387-977-11	Sequence 11, Appl
24	134	5.4	15	27	US-10-387-977-12	Sequence 12, Appl
25	134	5.4	15	27	US-10-387-977-13	Sequence 13, Appl
26	132.5	5.3	14	555	US-10-156-761-14592	Sequence 14592, A
27	132.5	5.3	15	691	US-10-369-493-23473	Sequence 23473, A
28	131.5	5.3	9	2122	US-09-813-214A-9	Sequence 9, Appli
29	131.5	5.3	12	2468	US-10-282-122A-66335	Sequence 66335, A
30	131.5	5.3	14	2468	US-10-246-330-4	Sequence 4, Appli
31	131	5.2	15	960	US-10-311-879-15	Sequence 15, Appl
32	130	5.2	15	25	US-10-387-977-81	Sequence 81, Appl
33	130	5.2	12	1946	US-10-282-122A-62947	Sequence 62947, A
34	129.5	5.2	12	1638	US-10-206-576-258	Sequence 258, App
35	129.5	5.2	12	1638	US-10-206-576-262	Sequence 262, App
36	129.5	5.2	12	1638	US-10-206-576-266	Sequence 266, App
37	129.5	5.2	12	1728	US-10-282-122A-56997	Sequence 5697, A
38	129	5.2	14	24	US-10-229-066-6	Sequence 6, Appli
39	129	5.2	12	2435	US-10-282-122A-47453	Sequence 47453, A
40	127	5.1	15	25	US-10-387-977-83	Sequence 83, Appl
41	127	5.1	15	25	US-10-387-977-84	Sequence 84, Appl
42	127	5.1	15	25	US-10-387-977-85	Sequence 85, Appl
43	126	5.1	15	1649	US-10-369-493-18460	Sequence 18460, A
44	123.5	5.0	15	491	US-10-387-977-100	Sequence 100, App
45	123.5	5.0	12	669	US-10-206-576-264	Sequence 264, App

```
                              ALIGNMENTS

RESULT 1
US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US2003015637A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric C.
;   APPLICANT: Bhogal, Peter S.
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;   FILE REFERENCE: Reynolds
;   CURRENT APPLICATION NUMBER: US/10/229,066
;   CURRENT FILING DATE: 2002-08-28
;   PRIOR APPLICATION NUMBER: US/09/066,330
;   PRIOR FILING DATE: 1998-09-15
;   PRIOR APPLICATION NUMBER: PN 6275
;   PRIOR FILING DATE: 1995-10-30
;   PRIOR APPLICATION NUMBER: PCT/AU96/00673
;   PRIOR FILING DATE: 1996-10-30
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 10
;   LENGTH: 1706
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match           83.5%;  Score 2070;  DB 14;  Length 1706;
Best Local Similarity 85.7%;  Pred. No. 3.9e-176;
Matches  390;  Conservative  18;  Mismatches  41;  Indels  6;  Gaps  4;

Qy     3  PNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASS 62
          PNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010
Db   954  PNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy    63  ASY-INFEGPQNPDNYLVTPELSLPNGGLTFWVCAQDANYASEHYAVVASSTGNDASNF 121
          ||:|  |||||||||||||||:||||||||:|||||||||||||||||||||||||||
Db  1011  ESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNF 1070
```

```
Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
        |||||:||||   |: ||||  :|||  : : |||||||||||||||||||||||||||
Db 1071 TNALLEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy  182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
        :|||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db 1129 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1188

Qy  242 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
        |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db 1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy  302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1249 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy  362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1309 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy  422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
        :||||||   |||||||:|||  ||:|||||  ||
Db 1369 KCVNVTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403
```

RESULT 2
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

```
Query Match          82.2%;  Score 2039.5;  DB 14;  Length 1732;
Best Local Similarity 84.5%;  Pred. No. 2.2e-173;
Matches  388;  Conservative  18;  Mismatches  44;  Indels  9;  Gaps  6;

Qy    1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
        ||||||||||||  |||||||||||||||||||||||   |||:||| |   : |:|
Db  969 GTPNPNPNPNPGTTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy   60 ASSASY-INFEGPQNPDNYLIVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
        ||: :  :|||| | ||||:||| ||||| |||| |||||||||||||||||||||||||
Db 1026 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy  119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
        |||:||||||||||||   |: :|||| |: ||| : |||||||||||||||   |: ||
Db 1086 SNFTNALLEETITAKG-VRSPAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy  179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG 238
        ||   :||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db 1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy  239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF 298
        |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy  299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
        |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db 1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy  359 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1324 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy  419 SPKECVNVTVDPVQPVQNPVQNLTGSAV--GQKVTLKWDAP 455
        |||:|||| |   |:|| |||:||  |    :||||  ||
Db 1384 SPKKCVNVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422
```

RESULT 3
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 5292820000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

```
Query Match          29.3%;   Score 726;  DB 15;  Length 419;
Best Local Similarity 75.3%;   Pred. No. 2.3e-56;
Matches  143;  Conservative  11;  Mismatches  30;  Indels  6;  Gaps  4;

Qy    3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
        ||||||||||||||||||||||||||||||||||   |||||||   |:||||||||
Db  235 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 291

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
        |:  :|||| | ||||:||| ||||| |||| ||||||||||||||||||||||||||||
Db  292 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 351

Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
        ||||||||||||   |: :|||| |: ||| : |||||||||||||||   |: ||
Db  352 TNALLEETITAKG-VRSPAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 409

Qy  182 DVEIKANGKR 191
        :|||||||||
Db  410 EVEIKANGKR 419
```

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 5292820000700
; CURRENT APPLICATION NUMBER: US/10/174,695

```
Db     92  RYDDFTE--AGKKYTFTMRRAGMGDGTDMEVED-------DSPASYTYVRDGTKIK 141
Qy    389  EGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKBCVNVTVD-PVQFNPVQNLTGSAVGGK 447
                |||||.||||||||||||||||||||:.:||.:  .||.:||||||||||:|
Db    142  EGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGGK 201

Qy    448  VTLKWDAPN 456
                |||||||||
Db    202  VTLKWDAPN 210
```

RESULT 6
US-10-174-695-4
; Sequence 4, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200070
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 4
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-4

Query Match   9.4%; Score 232.5; DB 15; Length 196;
Best Local Similarity 39.1%; Pred. No. 1.5e-12;
Matches   61; Conservative 13; Mismatches 63; Indels 19; Gaps 5;

```
Qy     3  PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
Db    44  PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYL 100
Qy    63  ASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFA 122
Db   101  DNSAKIDRNQRINVY-NTAEYAKTNAPI-----KVVGYADE------KTGTAAYNM- 145
Qy   123  NALLEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQL 158
Db   146  -KLSRRAKAVAKMLEKYGVSADRITIEWKGSSEQI 180
```

RESULT 7
US-10-387-977-18
; Sequence 18, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200030
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30

; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-5

Query Match   28.6%; Score 710.5; DB 15; Length 419;
Best Local Similarity 73.6%; Pred. No. 5.6e-55;
Matches 142; Conservative 12; Mismatches 32; Indels 7; Gaps 5;

```
Qy     1  GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
Db   232  GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 288
Qy    60  ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
Db   289  VYSESFGLGGIGVLITPDNYLITPALDLIPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 348
Qy   119  SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
Db   349  SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 406
Qy   179  NLDDVEIKANGKR 191
Db   407  DLDEVEIKANGKR 419
```

RESULT 5
US-10-174-695-6
; Sequence 6, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200070
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-6

Query Match   15.5%; Score 385; DB 15; Length 231;
Best Local Similarity 41.4%; Pred. No. 3.9e-26;
Matches 103; Conservative 31; Mismatches 69; Indels 46; Gaps 11;

```
Qy   213  IDADGDGQGWLCLSSGQLGWLTAHGGTNVV-ASFSWNGMALNPDNYLISKD--VTGATKV 269
Db     3  LDADHNTFGSVIPATGPL--FTGTASSNLYSANFEYL-IPANADPVVTQNIIVTGQGEV 59
Qy   270  KYYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEETPNGINKGGARFGLSTEADGAKPQS 329
Db    60  V------IPGGVY----------DYCITNPEPASG------KMWIAGDGGNQPA 91
Qy   330  VWIERTVDLPAGTKYV-AFRHYNCSDLNVILLDDIQFTMGGSPTPDYTYTVVRDGTKIK 388
```

```
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 18
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-18

Query Match        6.2%;  Score 153;  DB 15;  Length 29;
Best Local Similarity   100.0%;  Pred. No. 1.3e-06;
Matches  29;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
          |||||||||||||||||||||||||||||
Db      1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 8
US-10-387-977-21
;   Sequence 21, Application US/10387977
;   Publication No. US2004005276A1
;   GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 529282000301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 21
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-21

Query Match        6.2%;  Score 153;  DB 15;  Length 29;
Best Local Similarity   100.0%;  Pred. No. 1.3e-06;
Matches  29;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
          |||||||||||||||||||||||||||||
Db      1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 9
US-10-387-977-23
;   Sequence 23, Application US/10387977
;   Publication No. US2004005276A1
;   GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 529282000301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 23
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-23

Query Match        6.2%;  Score 153;  DB 15;  Length 29;
Best Local Similarity   100.0%;  Pred. No. 1.3e-06;
Matches  29;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
          |||||||||||||||||||||||||||||
Db      1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 10
US-10-387-977-24
;   Sequence 24, Application US/10387977
;   Publication No. US2004005276A1
;   GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 529282000301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 24
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-24

Query Match        6.2%;  Score 153;  DB 15;  Length 29;
Best Local Similarity   100.0%;  Pred. No. 1.3e-06;
Matches  29;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
          |||||||||||||||||||||||||||||
Db      1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 11
US-10-387-977-26
;   Sequence 26, Application US/10387977
;   Publication No. US2004005276A1
;   GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 529282000301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
```

```
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 26
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-26

Query Match             6.2%;   Score 153;  DB 15;  Length 29;
Best Local Similarity   100.0%;  Pred. No. 1.3e-06;
Matches   29;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy     375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
               |||||||||||||||||||||||||||||
Db       1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29


RESULT 12
US-10-387-977-17
;   Sequence 17, Application US/10387977
;   Publication No. US20040005276A1
;   GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 529282000301
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 17
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-17

Query Match             5.9%;   Score 147;  DB 15;  Length 29;
Best Local Similarity   96.6%;  Pred. No. 4.3e-06;
Matches   28;   Conservative    0;   Mismatches    0;   Indels    1;   Gaps    0;

Qy     375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
               |||||||||||||||||| |||||||||
Db       1 DYTYTVYRDGTKIKEGLTATTFEEDGVAT 29


RESULT 13
US-10-282-122A-55467
;   Sequence 55467, Application US/10282122A
;   Publication No. US20040029129A1
;   GENERAL INFORMATION:
;   APPLICANT: Wang, Liangsu
;   APPLICANT: Zamudio, Carlos
;   APPLICANT: Malone, Cheryl
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Kari
;   APPLICANT: Zyskind, Judith
;   APPLICANT: Wall, Daniel
;   APPLICANT: Trawick, John
;   APPLICANT: Carr, Grant
;   APPLICANT: Yamamoto, Robert
;   APPLICANT: Forsyth, R.
;   APPLICANT: Xu, H.
```

```
;   TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;   FILE REFERENCE: ELITRA.034A
;   CURRENT APPLICATION NUMBER: US/10/282,122A
;   CURRENT FILING DATE: 2003-02-20
;   PRIOR APPLICATION NUMBER: 60/191,078
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: 60/206,848
;   PRIOR FILING DATE: 2000-05-23
;   PRIOR APPLICATION NUMBER: 60/207,727
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: 60/230,335
;   PRIOR FILING DATE: 2000-09-06
;   PRIOR APPLICATION NUMBER: 60/230,347
;   PRIOR FILING DATE: 2000-09-09
;   PRIOR APPLICATION NUMBER: 60/242,578
;   PRIOR FILING DATE: 2000-10-23
;   PRIOR APPLICATION NUMBER: 60/253,625
;   PRIOR FILING DATE: 2000-11-27
;   PRIOR APPLICATION NUMBER: 60/257,931
;   PRIOR FILING DATE: 2000-12-22
;   PRIOR APPLICATION NUMBER: 60/267,636
;   PRIOR FILING DATE: 2001-02-09
;   PRIOR APPLICATION NUMBER: 60/269,308
;   PRIOR FILING DATE: 2001-02-16
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 78614
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 55467
;   LENGTH: 872
;   TYPE: PRT
;   ORGANISM: Enterobacter cloacae
US-10-282-122A-55467

Query Match             5.9%;   Score 147;   DB 12;   Length 872;
Best Local Similarity   21.3%;  Pred. No. 0.00064;
Matches   112;   Conservative   56;   Mismatches   205;   Indels   154;   Gaps    23;

Qy      10 NPGTTLSESFENGIPASWKTIDADGDGNWTTTPPGGTSFAGHNSAICASSASYINFE 69
            | |||.|| : .|:|:.|     ||.  :: :.    :  ::
Db     293 NTGTGTHNVTVDTGLPS--VAFNAISDDNVLNAVEKGQDLSVSGTSANLAEGTVVTVTLN 350

Qy      70 GPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASST--GNDASNFANALL 126
            |          ||:||.|| :|| :  |||:        : .   || ||  ||| |
Db     351 G---KNYAATTAAD----GTWSLTVPAADLAGLGDHYTLSATATNGVGNSVSNTANLLV 402

Qy     127 EEVLTAKT--------VVTAPEAIRGTRVQGTWTQKTVQLPAGTKYVAFRHFGCTDFFWI 178
            :   :        :       :  |         :: ||    |
Db     403 DTALPTVTINTVAGDNVINAAEVAAGQTISG----KVANAEAG--------------- 441

Qy     179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWICLSSGQLGWLTAHGG 238
            :  | |      :::: :   :::| |: :|:   : ::
Db     442 NTVVTIGGN---SYTATVQSDLTWSVNVPESVLTALGNGD--LTVSATV--TNGHGN 492

Qy     239 T-----NVVASFSWNGMALNPDNVLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGT 293
              |::  .    .:  | ||      ::. |    :
Db     493 TGAGBREIVIDASLPGLRIDT--VAGDDV--------------INSIEHGQNLIVTGSSDGL 538

Qy     294 NAGDFTVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGT----------- 342
            :      :    || :   :|:    :|  ||        ||||
Db     539 AAG--------TTLTVTVNGKTYAASVLADGTWSAAIPAADVGALAAGTVTVTAAGQSAA 590

Qy     343 -KYVAFRHVNCSDLNYILL-----------DDI------QFTMGGSPTPTD------- 375
            .|:||:|.|::|.||.|:                : ||     :::
Db     591 GNPVTISHDVTVDLSAVAISIDAIATDDVINAAEKGADIVLSGSTSNVBENQTVTITSGG 650

Qy     376 --YTYTVYRDG------------TKIKEG----LTETTFEEDGVATGNHEYCVEVKYTAGV 418
             :||:|:||                 :|:: ||:|   : :
Db     651 KTYTAKVDADGNWTATVPSADLAGLKDGDASVQVSVTNAHGNSASAGREYSVD----AT 705

Qy     419 SPKECVNVTVDPVQFNPVQNLTGSA---------------------VGQKVTLKWD 453
            : |:                                  :|: :| : |
Db     706 AP----TVTIDTVAGDNVINASEAAAGVAISGTTTAEVGQTVTVTLD 748
```

```
Qy  112  SSTGNDASNFANALL--EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQ--LPAGTKVVA 166
Db  123  ASSPEELTNIIDKVLMYEKATMPDKSYLEKVLLIAG-ADYSWNSQVGQPTIKYGMQYYY 180

Qy  167  FRHFGCTDFFWINLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL--C 224
Db  181  NQEHGYTDVY---------NYLKAPTGCYSHLNTGVSFANYT--AHGSETAWADPL 226

Qy  225  LSSGQLGWLT------AHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVND 277
Db  227  LTTSQLKALTNKDKYFLAIGNCCITAQFDY----VQP---CFGEVITRVKEKGAYAYIG 278

Qy  278  GFPGDH-------YAVMISKTGTNAGDFTVVF-EETPNGINKGGARFGLSTEA 322
Db  279  SSPNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVN------ 326

Qy  323  DGAKPQSVWIERTVDLPAGT---KYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT- 377
Db  327  -----SIMWAGNLAATHAGNIGNITHIGAHYY--WEAYHVLGDGSVMPYRAMPKTNTYTL 379

Qy  378  -----------------YTVYRDGTKIKEGLTETTFEEDGVAT----------GNHE 407
Db  380  PASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANAS----GVATVSMTKQITENGNYD 435

Qy  408  YCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
Db  436  VVITRSNYLPVIKQIQVG---EPSPYQPVSNLTATTQQKVTLKWEAPS 481
```

Search completed: May 18, 2004, 11:52:20
Job time : 35.3211 secs

RESULT 14
US-10-387-977-25
; Sequence 25, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-25

Query Match   5.8%; Score 145; DB 15; Length 29;
Best Local Similarity   96.6%; Pred. No. 6.6e-06;
Matches   28; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy 375 DYTYTVVRDGTKIKEGLTETTFEEDGVAT 403
         |||||||||||| ||||||||||||||||
Db   1 DYTYTVVRDGTKIKWGLTETTFEEDGVAT 29

RESULT 15
US-10-387-977-101
; Sequence 101, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-101

Query Match   5.8%; Score 143; DB 15; Length 509;
Best Local Similarity   20.0%; Pred. No. 0.00066;
Matches   94; Conservative   66; Mismatches   189; Indels   120; Gaps   21;

Qy   54 HNSAICASSAS-YINFEGPQNPDNYLVTPELSLPNGGTLTFWCAQDANYASBHYAV-YA 111
Db   67 YNDGLAASAAPVFLALVG----DTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMS 122
```

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:        May 18, 2004, 11:33:39 ; Search time 47.0287 Seconds
                                         (without alignments)
                                         2739.638 Million cell updates/sec

Title:         US-08-570-311-20
Perfect score: 2480
Sequence:      1 GTPNPNPNPGTTTLSESF.......QNLTGSAVGQKVTLKWDAPN 456

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:    1586107


Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     A_Geneseq.29Jan04:*
               1: geneseqp1980s:*
               2: geneseqp1990s:*
               3: geneseqp2000s:*
               4: geneseqp2001s:*
               5: geneseqp2002s:*
               6: geneseqp2003as:*
               7: geneseqp2003bs:*
               8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                           SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2480	100.0	456	2	AAR96023	Aar96023 P. gingiv
2	2480	100.0	456	2	AAR96022	Aar96022 P. gingiv
3	2480	100.0	456	2	AAW69491	Aaw69491 Haemagglu
4	2480	100.0	2628	2	AAW69490	Aaw69490 Haemagglu
5	2480	100.0	2628	2	AAR96030	Aar96030 P. gingiv
6	2480	100.0	450	2	AAW69488	Aaw69488 Haemagglu
7	2436	98.2	450	2	AAR96021	Aar96021 P. gingiv
8	2436	98.2	439	2	AAW69489	Aaw69489 Haemagglu
9	2351	94.8	439	2	AAR96024	Aar96024 P. gingiv
10	2351	94.8	1687	2	AAW69492	Aaw69492 Haemagglu
11	2082	84.0	1687	2	AAR96033	Aar96033 P. gingiv
12	2082	84.0	1704	2	AAW69495	Aaw69495 Haemagglu
13	2082	84.0	1704	2	AAR70188	Aar70188 Arg-gingi
14	2082	84.0	1704	2	AAW34843	Aaw34843 Arg-gingi
15	2082	84.0	1704	2	AAY67396	Aay67396 Arg-gingi
16	2082	84.0	1704	4	AAU08938	Aau08938 P. gingiv
17	2070	83.5	1706	2	AAW24786	Aaw24786 PrtR anti
18	2066	83.3	1087	2	AAR96028	Aar96028 P. gingiv
19	2066	83.3	1087	2	AAW69486	Aaw69486 Haemagglu
20	2066	83.3	1358	2	AAR96032	Aar96032 P. gingiv
21	2066	83.3	1358	2	AAW69494	Aaw69494 Haemagglu
22	2034.5	82.0	1732	2	AAR96029	Aar96029 P. gingiv
23	2034.5	82.0	1732	2	AAW24787	Aaw24787 PrtK anti
24	2034.5	82.0	1732	2	AAW69487	Aaw69487 Haemagglu
25	726	29.3	419	4	AAU03572	Aau03572 P. gingiv

26	720	29.0	135	6	ABP55081	Abp55081 Porphyrom
27	710.5	28.6	419	4	AAU03574	Aau03574 P. gingiv
28	704	28.4	134	4	AAB49217	Aab49217 Peptide u
29	682	27.5	921	2	AAY34522	Aay34522 Porphyrym
30	682	27.5	922	2	AAY34521	Aay34521 Porphyrym
31	682	27.5	925	2	AAY34520	Aay34520 Porphyrym
32	682	27.5	938	2	AAY34392	Aay34392 Porphyrym
33	565.5	22.8	377	2	AAY34359	Aay34359 Porphyrym
34	561.5	22.6	312	2	AAY34484	Aay34484 Porphyrym
35	546.5	22.0	497	2	AAR96025	Aar96025 P. gingiv
36	546.5	22.0	497	2	AAW69483	Aaw69483 Haemagglu
37	415	16.7	970	2	AAR72458	Aar72458 Porphyrom
38	385	15.5	231	4	AAU03575	Aau03575 P. gingiv
39	267	10.8	991	2	AAR77313	Aar77313 Porphyrom
40	236	9.5	49	2	AAW34805	Aaw34805 Arg-speci
41	233	9.4	293	2	AAY34483	Aay34483 Porphyrym
42	233	9.4	299	2	AAY34358	Aay34358 Porphyrym
43	232.5	9.4	196	4	AAU03573	Aau03573 P. gingiv
44	189	7.6	46	2	AAW34798	Aaw34798 Arg-speci
45	153	6.2	29	2	AAW83085	Aaw83085 Peptide f

```
                                             ALIGNMENTS

RESULT 1
AAR96023
ID   AAR96023 standard; protein; 456 AA.
XX
AC   AAR96023;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep3 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep3.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A,  Tunwasorn S,  Lepine G,  Han N,  Lantz M,  Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30647.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 110-112; 153pp; English.
XX
CC   HArep3 (AAR96023) is the product of the HArep3 repeat unit (AAT30647) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep3 and other hagA repeat
CC   unit products (see also AAR96021-22 and AAR96024) can be obtd. from
CC   transformed host cells and used as vaccines to protect humans or animals
CC   against periodontal disease. Expression in Salmonella cells allows prodn.
CC   of live vaccine. HArep1-4 can also be used to detect the presence of anti
CC   -P. gingivalis antibodies and to raise monoclonal antibodies for
CC   diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 456 AA;
```

```
CC   HArep2 (AAR96022) is the product of the HArep2 repeat unit (AAT30646) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep2 and other hagA repeat
CC   unit products (see also AAR96021 and AAR96023-24) can be obtd. from
CC   transformed host cells and used as vaccines to protect humans or animals
CC   against periodontal disease. Expression in Salmonella cells allows prodn.
CC   of live vaccine. HArep1-4 can also be used to detect the presence of anti
CC   -P. gingivalis antibodies and to raise monoclonal antibodies for
CC   diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 456 AA;

Query Match             100.0%;  Score 2480;  DB 2;  Length 456;
Best Local Similarity   100.0%;  Pred. No. 7.5e-191;
Matches   456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy  301 VFEEETPNGINKGGARPGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEEETPNGINKGGARPGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456

RESULT 3
AAW69491
ID   AAW69491 standard; protein; 456 AA.
XX
AC   AAW69491;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagA, HArep3.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
```

```
Query Match             100.0%;  Score 2480;  DB 2;  Length 456;
Best Local Similarity   100.0%;  Pred. No. 7.5e-191;
Matches   456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy  301 VFEEETPNGINKGGARPGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEEETPNGINKGGARPGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456

RESULT 2
AAR96022
ID   AAR96022 standard; protein; 456 AA.
XX
AC   AAR96022;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep2 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep2.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30646.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 107-108; 153pp; English.
XX
```

PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV5877.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 127-132; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 456 AA;

Query Match   100.0%;   Score 2480;   DB 2;   Length 456;
Best Local Similarity   100.0%;   Pred. No. 7.5e-191;
Matches 456; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy   421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
           ||||||||||||||||||||||||||||||||||||
Db   421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456

RESULT 5
AAR96030
ID   AAR96030 standard; protein; 2628 AA.
XX

PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58878.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 133-138; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 456 AA;

Query Match   100.0%;   Score 2480;   DB 2;   Length 456;
Best Local Similarity   100.0%;   Pred. No. 7.5e-191;
Matches 456; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy   421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
           ||||||||||||||||||||||||||||||||||||
Db   421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456

RESULT 4
AAW69490
ID   AAW69490 standard; protein; 456 AA.
XX
AC   AAW69490;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagA, HArep2.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX

```
AC   AAR96030;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagA.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
FH   Key             Location/Qualifiers
FT   Peptide         5..21
FT                   /label= Sig_peptide
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30654.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 6; Page 93-101; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagA (AAR96030) was identified as the
CC   product of the hagA gene (AAT30654) isolated as an EcoRV fragment of
CC   genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
CC   24), can be obtd. from transformed host cells and used as a vaccine to
CC   protect humans or animals against periodontal disease. Expression in
CC   Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
CC   also be used to detect the presence of anti-P. gingivalis antibodies and
CC   to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-
CC   2003 to standardise OS field)
XX
SQ   Sequence 2628 AA;

Query Match        100.0%;     Score 2480;  DB 2;    Length 2628;
Best Local Similarity  100.0%;   Pred. No. 9e-190;
Matches  456;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy       1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA     60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     950 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA   1009

Qy      61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN    120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1010 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN   1069

Qy     121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL    180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1070 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL   1129

Qy     181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTN    240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1130 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTN   1189

Qy     241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAMISKTGTNAGDFTV    300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1190 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAMISKTGTNAGDFTV   1249

Qy     301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL    360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1250 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL   1309

Qy     361 DDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP    420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1310 DDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP   1369

Qy     421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN    456
           ||||||||||||||||||||||||||||||||||||
Db    1370 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN   1405


RESULT 6
AAW69488
ID   AAW69488  standard; protein; 2628 AA.
XX
AC   AAW69488;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagA.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58875.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 91-110; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 2628 AA;

Query Match        100.0%;     Score 2480;  DB 2;    Length 2628;
Best Local Similarity  100.0%;   Pred. No. 9e-190;
Matches  456;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy       1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA     60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     950 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA   1009

Qy      61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN    120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1010 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN   1069

Qy     121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL    180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1070 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL   1129
```

```
Qy    181  DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGTN  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130  DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGTN  1189

Qy    241  VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190  VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV  1249

Qy    301  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL  360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL  1309

Qy    361  DDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP  420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310  DDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP  1369

Qy    421  KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  456
           ||||||||||||||||||||||||||||||||||||
Db   1370  KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  1405
```

```
RESULT 7
AAR96021
AAR96021
ID   AAR96021 standard; protein; 450 AA.
XX
AC   AAR96021;
XX
DT   13-JUN-1996.
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep1 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep1.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30645.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 103-104; 153pp; English.
XX
CC   HArep1 (AAR96021) is the product of the HArep1 repeat unit (AAT30645) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep1 and other hagA repeat
CC   unit products (see also AAR96022-24) can be obtd. from transformed host
CC   cells and used as vaccines to protect humans or animals against
CC   periodontal disease. Expression in Salmonella cells allows prodn. of live
CC   vaccine. HArep1-4 can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 450 AA;

Query Match             98.2%;  Score 2436;  DB 2;  Length 450;
Best Local Similarity   99.8%;  Pred. No. 2.6e-187;
Matches  449;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy      7  PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI  66
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI  60

Qy     67  NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL  126
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL  120

Qy    127  EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK  186
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK  180

Qy    187  ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGTNVVASFS  246
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGTNVVASFS  240

Qy    247  WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP  306
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP  300

Qy    307  NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT  366
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT  360

Qy    367  MGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV  426
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  MGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV  420

Qy    427  TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  456
           ||||||||||||||||||||||||||||||
Db    421  TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  450
```

```
RESULT 8
AAW69489
AAW69489
ID   AAW69489 standard; protein; 450 AA.
XX
AC   AAW69489;
XX
DT   22-DEC-1998  (first entry)
XX
DE   Haemagglutinin protein hagA, HArep1.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58876.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s).
XX
PS   Claim 1; Col 121-126; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
```

```
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 450 AA;

Query Match        98.2%;     Score 2436;  DB 2;  Length 450;
Best Local Similarity 99.8%;  Pred. No. 2.6e-187;
Matches 449;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy    7 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 66
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60

Qy   67 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120

Qy  127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 186
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180

Qy  187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFS 246
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240

Qy  247 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300

Qy  307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360

Qy  367 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 426
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420

Qy  427 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
         ||||||||||||||||||||||||||||||
Db  421 TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 450

RESULT 9
AAR96024
ID   AAR96024 standard; protein; 439 AA.
XX
AC   AAR96024;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep4 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep4.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995; 95WO-US016108.
XX
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
```

```
DR   N-PSDB; AAT30648.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 114-115; 153pp; English.
XX
CC   HArep4 (AAR96024) is the product of the HArep4 repeat unit (AAT30648) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep4 and other hagA repeat
CC   unit products (see also AAR96021-23) can be obtd. from transformed host
CC   cells and used as vaccines to protect humans or animals against
CC   periodontal disease. Expression in Salmonella cells allows prodn. of live
CC   vaccine. HArep1-4 can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 439 AA;

Query Match        94.8%;     Score 2351;  DB 2;  Length 439;
Best Local Similarity 98.2%;  Pred. No. 1.7e-180;
Matches 431;  Conservative 3;  Mismatches 5;  Indels 0;  Gaps 0;

Qy    1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
         |  |||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYAVYASSTGNDASN 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYAVYASSTGNDASN 120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
         |:||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db  181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421 KECVNVTVDPVQFNPVQNL 439
         |:|||||:::| |||||||
Db  421 KVCVNVTINPTQFNPVQNL 439

RESULT 10
AAW69492
ID   AAW69492 standard; protein; 439 AA.
XX
AC   AAW69492;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagA, HArep4.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
```

```
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58879.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(B);
XX
PS   Claim 1; Col 139-144; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the haGA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 439 AA;

Query Match             94.8%;  Score 2351;  DB 2;  Length 439;
Best Local Similarity   98.2%;  Pred. No. 1.7e-180;
Matches   431;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    1  GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICA  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICV  60

Qy   61  SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
         |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   61  SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy  121  FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181  DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
         | ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db  181  DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240

Qy  241  VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGPPGDHYAVMISKTGTNAGDFTV 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGPPGDHYAVMISKTGTNAGDFTV 300

Qy  301  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361  DDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  DDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421  KECVNVTVDPVQFNPVQNL 439
         | |||||:::||||||||||
Db  421  KVCVNVTINPTQFNPVQNL 439

RESULT 11
AAR96033
ID   AAR96033 standard; protein; 1687 AA.
XX
AC   AAR96033;
XX
DT   16-OCT-2003 (revised)
```

```
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagE.
XX
KW   Haemagglutinin; hagE; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain FDC381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30656.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 138-143; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagE (AAR96033) was identified as the
CC   product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
CC   The haemagglutinin can be obtd. from transformed host cells and used as a
CC   vaccine to protect humans or animals against periodontal disease.
CC   Expression in salmonella cells allows prodn. of a live vaccine. The
CC   haemagglutinin can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1687 AA;

Query Match             84.0%;  Score 2082;  DB 2;  Length 1687;
Best Local Similarity   86.2%;  Pred. No. 5.2e-158;
Matches   392;  Conservative   17;  Mismatches   40;  Indels    6;  Gaps    4;

Qy    3  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASS   62
         |||||||||||||||||||||||||||||||||||: |     :||::|:  :|
Db  935  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS  991

Qy   63  ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          :  :  ||||| ||||: |  ||||| |||||||||||||||||||||||||||||||
Db  992  ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy  122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          ||||| :::|| : |:| ||||: :|||| :||| |||||||||||||| :||:|:|:||
Db 1052  TNALLEETTIAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy  182  DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 1110  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1169

Qy  242  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGPPGDHYAVMISKTGTNAGDFTVV 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1170  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGPPGDHYAVMISKTGTNAGDFTVV 1229

Qy  302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
         |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db 1230  FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1289

Qy  362  DIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1290  DIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1349

Qy  422  ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

Db   1350 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1384

RESULT 12
AAW69495
ID   AAW69495 standard; protein; 1687 AA.
XX
AC   AAW69495;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagE.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagE.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58881.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s).
XX
PS   Claim 1; Col 167-182; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagE haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1687 AA;

Query Match   84.0%;   Score 2082;   DB 2;   Length 1687;
Best Local Similarity   86.2%;   Pred. No. 5.2e-158;
Matches 392;   Conservative 17;   Mismatches 40;   Indels 6;   Gaps 4;

Qy   3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASS 62
            ||||||||||||||||||||||||||||||||||||: | |:|| | ||||||
Db   935 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 991

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
            |:|: |||||||| ||||| |||:|||||||||||||||||||||||||||||||||
Db   992 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
            | |||| ||| |:||||:||| :|||||||||||||||||||||| ||:|: |||:
Db   1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGTNV 241
            ||||||||||||||||||||||||||||||||||||||||| :||||||||||||||
Db   1110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1169

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1229

PT Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
PT useful for protecting animals and humans from gingivalis and periodontal
PT diseases.
XX
PS Disclosure; Page 68-73; 95pp; English.
XX
CC The present sequence represents an arginine-specific protease of
CC Porphyromonas gingivalis. The following peptides, derived from Arg- and
CC Lys-specific high molecular weight proteases, offer protection against
CC infection: YTYTVYRDGK IKEGLTATTF DDGVATGNHE YCVEKYTAGS VSPKVC (I);
CC YTPVEEKQNG RMIVIVAKKY (II); QLPFIFDVAC VNGDFLFSMP CFAEALMRAQ (III);
CC GEPNYYQPVS NLTATTQGQK VTLKWDAPSTK (IV); GNHEYCVEVK YTAGVSPKVC KDVTV (V);
CC RMFMNYEPGR YTPVEEKQNG (VI), TFAGFEDYK RMFMNYEPGR (VII); DYTYTVYRDG
CC TKIKEGLTAT TFEEDGVATG NMEYCVCVKY TAGVSPKVC (VIII); YTYTVYRDGT KIKEGLTATTF
CC EEDG (IX); RDGTKIKEGL TATTFEEDGV ATGN (X); KIKEGLTATT FEEDGVATGN HEY (XI)
CC ; KWDAPNGTPN PNPN PGTTTLSE (XII); and YTPVEEKENG RMIVIVAKKY
CC (XIII). They are used in vaccines to protect animals, including humans,
CC from gingivitis and/or periodontal diseases
XX
SQ Sequence 1704 AA;

```
Query Match            84.0%;  Score 2082;  DB 2;  Length 1704;
Best Local Similarity  86.2%;  Pred. No. 5.2e-158;
Matches  392;  Conservative  17;  Mismatches  40;  Indels  6;  Gaps  4;

Qy      3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASS     62
          ||||||||||||||||||||||||||||||||||||||       ||||  :||  ||
Db    952 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy     63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF  121
          :|  ||:|||||||||||||:||||||||||||||||||||||||||
Db   1009 ESFGLGLGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy    122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD  181
          :||||::|:||| :|| ||||| ||||||:|||||||||||||||||  :|| |:||||
Db   1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy    182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV  241
          :||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy    242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy    302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
          ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db   1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy    362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy    422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  456
          |||||:   |||||::|| |  :  ::|||| ||:
Db   1367 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

RESULT 14
AAW34843
ID   AAW34843 standard; protein; 1704 AA.
XX
AC   AAW34843;
XX
DT   03-JUN-1998 (first entry)
XX
DE   Arg-gingipain high molecular weight prepolyprotein sequence.
XX
KW   Arg-specific gingipain protease; gingivalis; periodontal disease;
KW   vaccine; infection.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key   Location/Qualifiers
FT   Protein   1..227
FT   /note= "precursor protein"
XX
PN   WO9734629-A1.
XX
PD   25-SEP-1997.
XX
PF   21-MAR-1997; 97WO-US004635.
XX
PR   22-MAR-1996; 96US-0013945P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
PA   (MORE-) MOREHOUSE SCHOOL MEDICINE.
XX
PI   Potempa J, Travis J, Genco C;
XX
DR   WPI; 1997-479993/44.
DR   N-PSDB; AAT93872.
XX

RESULT 15
AAY67396
ID   AAY67396 standard; protein; 1704 AA.
XX
AC   AAY67396;
XX
DT   25-APR-2000 (first entry)
XX
DE   Arg-gingipain-2 amino acid sequence.
XX
KW   Arginine specific proteinase; Arg-gingipain; gingipain; gingipain-2; haemagglutinin;
KW   immunogenic component; vaccine; inflammatory response; tissue damage;
KW   periodontal disease.
XX
OS   Porphyromonas gingivalis.

```
Qy       422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
             |||||::||  |||||| |   |||  |||||
Db      1367 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

Search completed: May 18, 2004, 11:42:43
Job time : 48.0287 secs

```
XX
FH   Key              Location/Qualifiers
FT   Region           229..719
FT                    /note= "Amino acids 229-719 are specifically claimed"
FT   Region           720..1185
FT                    /note= "Amino acids 720-1185 are specifically claimed"
XX
PN   US6017532-A.
XX
PD   25-JAN-2000.
XX
PF   08-NOV-1994;   94US-00336308.
XX
PR   10-SEP-1993;   93US-00119361.
PR   24-JUN-1994;   94US-0026544l.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Potempa JS, Travis J;
XX
DR   WPI; 2000-136659/12.
DR   N-PSDB; AAZ60181.
XX
PT   New Porphyromonas gingivalis arginine-specific protease preparation
PT   useful for preparing vaccines against periodontal disease and for
PT   screening for Arg-gingipain inhibitors.
XX
PS   Claim 1; Col 29-42; 55pp; English.
XX
CC   This sequence represents a Porphyromonas gingivalis arginine-specific
CC   proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
CC   Gingipain-2 consists of a 50kD protease component non-covalently
CC   associated with a 44kD haemagglutinin component. The proteinase is
CC   stimulated by glycine containing peptides and glycine analogues. It is
CC   inhibited by cysteine protease group specific inhibitors. The protease
CC   preparation can be used in immunogenic compositions and vaccines against
CC   periodontal disease. It can also be used to screen for agents that
CC   inflammatory response and tissue damage caused by P. gingivalis in
CC   modulate Arg-gingipain proteinase activity inhibitors
XX
SQ   Sequence 1704 AA;
```

```
Query Match        84.0%;   Score 2082; DB 3;   Length 1704;
Best Local Similarity   86.2%;   Pred. No. 5.2e-158;
Matches 392;   Conservative  17;   Mismatches  40;   Indels   6;   Gaps   4;
```

```
Qy       3 PNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
           |||||||||||||||||||||||||||||:|  |  |:|| || ||    |||
Db     952 PNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy      63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
           ||: :                  ||||||||||||||||||||||||||||||||||||
Db    1009 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy     122 ANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDPFWINLD 181
           ||||||   || |||| |   :||||||||||||||||||||||||||| :|:|
Db    1069 TNALLEETTTAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy     182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
           :|:|||||||||||||||||||||||||||||||||||||| |||||||| |||||||
Db    1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy     242 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
           |||||||||||||||||||:||||||||||||||||||||||||||:||||||||||||
Db    1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy     302 FEETPNGINKGGARPGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
           ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    1247 FEETPNGINKGGARPGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy     362 DIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1307 DIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366
```

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 18, 2004, 11:37:00 ; Search time 11.906 Seconds
                                    (without alignments)
                                    3684.135 Million cell updates/sec

Title:          US-08-570-311-20
Perfect score:  2480
Sequence:       1 GTPNPNPNPGTTLSESF........QNLTGSAVGQKVTLKWDAPN 456

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    PIR_78:*
              1:  pir1:*
              2:  pir2:*
              3:  pir3:*
              4:  pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description			
1	2480	100.0	2628	2	T28651	hemagglutinin A -			118.5	4.8	4199	2	S76412	hypothetical prote
2	2082	84.0	1704	2	A55426	gingipain R (EC 3.			
3	2059	83.0	1526	2	S49763	gingipain R (EC 3.			
4	2034.5	82.0	1732	2	T30836	lysine-specific cy			
5	266	10.7	991	2	I40029	arginyl endopeptid			
6	139.5	5.6	1052	2	AF2959	conserved hypothet			
7	139.5	5.6	1341	2	H98323	hypothetical prote			
8	132.5	5.3	691	2	B75622	hypothetical prote			
9	131.5	5.3	2468	2	A83412	hypothetical prote			
10	128.5	5.2	1248	2	C89874	autolysin limporte			
11	128	5.2	4936	2	AH2515	hypothetical prote			
12	127.5	5.1	1684	2	S10789	amylase A-180 - al			
13	127	5.1	713	2	B75489	hypothetical prote			
14	126	5.1	1441	2	A86685	prophage pil prote			
15	126	5.1	1649	2	C86822	hypothetical prote			
16	124.5	5.0	1904	2	T13256	tail-host specific			
17	124.5	5.0	5291	2	F90696	hypothetical prote			
18	123.5	5.0	5188	2	B85547	probable RTX famil			
19	122.5	4.9	1274	2	T10729	transferrin-like p			
20	121.5	4.9	465	2	A47023	S-layer protein -			
21	121.5	4.9	3624	2	AD0835	large repetitive p			
22	121	4.9	1090	2	S59077	cellulose 1,4-beta			
23	120	4.8	908	2	AE2254	hypothetical prote			
24	120	4.8	2783	2	T34416	hypothetical prote			
25	119	4.8	715	2	JC4908	alkaline serine pr			
26	119	4.8	1873	2	T30944	surface protein pr			
27	118.5	4.8	926	2	D86897	hypothetical prote			
28	118.5	4.8	1385	2	T18213	parasporal crystal			
29	118.5	4.8	2817	2	B97033	uncharacterized pr			

```
                                                       30  118.5  4.8  4199  2  S76412  hypothetical prote
                                                       31  118    4.8  1461  2  E90696  hypothetical prote
                                                       32  118    4.8  1461  2  A85547  hypothetical prote
                                                       33  117.5  4.7  1034  2  T30551  beta-galactosidase
                                                       34  116.5  4.7  1635  2  AI0452  hemolysin limporte
                                                       35  115.5  4.7   702  2  S48753  major surface prot
                                                       36  115.5  4.7   875  2  AF0472  probable outer mem
                                                       37  115.5  4.7  1118  2  A49724  protein-tyrosine-p
                                                       38  115.5  4.7  1282  2  JC4393  microbial collagen
                                                       39  115.5  4.7  1939  2  D97316  probable s-layer p
                                                       40  115    4.6   821  2  AD1507  probable secreted
                                                       41  115    4.6  1651  2  JC1340  outer membrane pro
                                                       42  115    4.6  3029  2  S76109  hypothetical prote
                                                       43  114    4.6   872  2  S49541  cellulase - Cellul
                                                       44  113    4.6  1345  2  H90975  hypothetical prote
                                                       45  112.5  4.5   938  2  AF1772  internalin-like pr

                                           ALIGNMENTS

RESULT 1
T28651
hemagglutinin A - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C;Accession: T28651
R;Han, N.; Whitlock, J.; Progulske-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A;Title: The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381 contains four
A;Reference number: Z20494; MUID:97047672; PMID:8926061
A;Accession: T28651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2628 <HAN>
A;Cross-references: EMBL:U41807; NID:g1552410; PID:g1469916; PIDN:AAB17128.1
C;Genetics:
A;Gene: hagA

Query Match             100.0%;  Score 2480;  DB 2;  Length 2628;
Best Local Similarity   100.0%;  Pred. No. 1.4e-162;
Matches  456;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNWTTTPPGGTSFAGHNSAICA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   950 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNWTTTPPGGTSFAGHNSAICA 1009

Qy    61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDASN 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1010 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDASN 1069

Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1070 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 1129

Qy   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1130 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 1189

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1190 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 1249

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1250 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1309

Qy   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1310 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1369

Qy   421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
           ||||||||||||||||||||||||||||||||||||
```

Db   1370 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 1405

RESULT 2
A55426

gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
N;Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP; R
C;Species: Porphyromonas gingivalis
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
C;Accession: A55426; D53113
R;Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr, F
J. Biol. Chem. 270, 1007-1010, 1995
A;Title: Molecular cloning and structural characterization of the Arg-gingipain proteina
A;Reference number: A55426; MUID:95138080; PMID:7836351
A;Accession: A55426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1704 <PAV>
A;Cross-references: GB:U15282; NID:g557067; PIDN:AAA69539.1; PID:g557068
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: D53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 228-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141694)
C;Keywords: cysteine proteinase; hydrolase

Query Match   84.0%;   Score 2082;   DB 2;   Length 1704;
Best Local Similarity   86.2%;   Pred. No. 2.2e-135;
Matches   392;   Conservative   17;   Mismatches   40;   Indels   6;   Gaps   4;

Qy   3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
        |||||||||||||||||||||||||||||||||||||||||::| |||||||||:||||
Db   952 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
        |:: | |||||||||||||||||||||||||||||||||||||||||||||
Db   1009 ESFGLGGIGVLTPDNYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
        ||||| |:|| :||||| |:|||||||:|:|||||||||||||||:||:|
Db   1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTMFYIDLD 1126

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
        :|||||||||||||||||||||||||||||||||||||| | ||| ||||||||||||
Db   1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
        |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307 DIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy   422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
        ||||||:::|:||| | | ||:
Db   1367 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401

RESULT 3
S49763

gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis (fragment)
C;Species: Porphyromonas gingivalis
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997

C;Accession: S49763
R;Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.
submitted to the EMBL Data Library, November 1994
A;Description: Cloning, sequence analysis and expression in Escherichia coli of prpR1 of
A;Reference number: S49763
A;Accession: S49763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <ADU>
A;Cross-references: EMBL:X82680
C;Genetics:
A;Gene: prpR1
C;Keywords: cysteine proteinase; hydrolase

Query Match   83.0%;   Score 2059;   DB 2;   Length 1526;
Best Local Similarity   85.5%;   Pred. No. 7.4e-134;
Matches   389;   Conservative   17;   Mismatches   43;   Indels   6;   Gaps   4;

Qy   3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
        |||||||||||||||||||||||||||||||||||||||||::| |||||||||:||||
Db   954 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
        |:: | |||||||||||||||||||||||||||||||||||||||||||||
Db   1011 ESFGLGGIGVLTPDNYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNF 1070

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
        ||||| |:|| :||||| |:|||||||:|:||||||||||||||||:||:|
Db   1071 TNALLEETITAKG-VRSPEAIRG-RIQSTWRQKTVDLPAGTKYVAFRHFQSTMFYIDLD 1128

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
        :|||||||||||||||||||||||||||||||||||||| | ||| ||||||||||||
Db   1129 EVEIKANGKRADFTETFESSTHGEATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1188

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
        |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1249 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy   362 DIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1309 DIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy   422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
        ||||||:::|:||| | | ||:
Db   1369 KCVNVTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 4
T30836

lysine-specific cysteine proteinase porphypain (EC 3.4.22.-) - Porphyromonas gingivalis
N;Alternate names: lysine-specific cysteine proteinase 1, 60K
C;Species: Porphyromonas gingivalis
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30836; T30837; T30526; A53113
R;Barkocy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulske-Fox, A.; Lantz,
J. Bacteriol. 178, 2734-2741, 1996
A;Title: Analysis of the prtP gene encoding porphypain, a cysteine proteinase of Porphyr
A;Reference number: Z20895; MUID:96213011; PMID:8631659
A;Accession: T30836
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1732 <BAR>
A;Cross-references: EMBL:U42210; NID:g1314325; PID:g1314326; PIDN:AAB06565.1
R;Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z20896
A;Accession: T30837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA

A;Residues: 1-795,'I',797-1389,'N',1391-1478,'Y',1480-1732 <SLA>
A;Cross-references: EMBL:U75366; NID:g2182811; PID:g2182812; PIDN:AAB60809.1
R;Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A;Title: IS195, an insertion sequence-like element associated with protease genes in Por
A;Reference number: 220844; MUID:98298016; PMID:9633563
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1350,'N',1352-1363,'Y',1365-1447,'H',1449-1732 <LEW>
A;Cross-references: EMBL:AF017055; NID:g2738802; PID:g2738803; PIDN:AAC26523.1
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: A53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 229-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141690)
C;Genetics:
A;Gene: prtP; prtK
C;Keywords: cysteine proteinase; hydrolase

Query Match   82.0%;   Score 2034.5; DB 2;   Length 1732;
Best Local Similarity   84.3%; Pred. No. 4.3e-132;
Matches   387; Conservative   19; Mismatches   44; Indels   9; Gaps   6;

Qy   1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
          ||||||||||||| :|||||||||||||||||||| :|||||:||||||||
Db   969 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy   60 ASSASY-INFEGPQNPDNYLVTPELSLPNGGLTFWVCAQDANYASEHYAVYASSTGNDA 118
          :||| |||||:|||||||||||| :|:||||||||||||||||||||||||||||
Db   1026 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy   119 SNFPANALLEEVLTAKTVVTAPEAIRGTRVQGTWVQKTVQLPAGTKYVAFRHFGCTDFPWI 178
          ||| |||||||||||:|: |:|||: |||||||||||||||||||||||||||
Db   1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy   179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGWLCLSSGQLGWLTAHGG 238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGWLCLSSGQLDWLTAHGG 1203

Qy   239 TNVVASFSWNGWMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
          :||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 SNVVSSFSWNGWMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy   299 TVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
          ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db   1264 TVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy   359 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy   419 SPKECVNVTVDPVQFNVQNLTGSAV--GQKVTLKWDAP 455
          ||||||:||||||||||||||||: ||||||||||||
Db   1384 SPKKCVDVTVNSTQFNVQNLTAEQAPNSMDAILKWNAP 1422

RESULT 5
I40229
arginyl endopeptidase - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40229
R;Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A;Title: Structural characterization of argingipain, a novel arginine-specific cysteine
A;Reference number: I40229; MUID:95168884; PMID:7864651
A;Accession: I40229

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-991 <RES>
A;Cross-references: GB:D26470; NID:g927644; PIDN:BAA05484.1; PID:g927645

Query Match   10.7%;   Score 266; DB 2;   Length 991;
Best Local Similarity   25.9%; Pred. No. 2e-10;
Matches   114; Conservative   50; Mismatches   164; Indels   112; Gaps   19;

Qy   30 TIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYINFEGPQNPDNYLVTPELSLPNGG 89
          ::: |:   :   :::
Db   595 SVNVSCDYNGAIATISANGKMF---GSAVVENGTATINLTC------------LTNES 637

Qy   90 TLTFWVCAQDANYASEHYAVYASSTG--NDASNFANALLEEVLTAKTVVTAPEAIRGTRV 147
          ||||:   :: : .: |::: |:||||:|
Db   638 TLTLTV---VGYNKETVIKTINTNGEPNYQPVSN---LTATT-----------OGQKV 679

Qy   148 QGTWVQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV-EIKANGKRADFTETFESSTHGEA 206
          ::   :|:   ::: |:   :|:   :|
Db   680 TLKWDAPSTKTNATTN-TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVVNDGS- 737

Qy   207 PAEWTTIDADGDGGQWLCLSSGQLGWLTAHGGTNVVASFSWN-----GMALNPDNYLISK 261
          :: |:   :::|::
Db   738 -GYQILLDADHDQYGQVIPSDTHTLWPNCSVPANLFAPFEYTVPENADPSCSPTNMIM-- 794

Qy   262 DVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTE 321
          |:||::   :||: |::
Db   795 DGTASVNI-----------------PAGTV------------DFAI----------- 811

Qy   322 ADGAKPQS---VWI------ERTVDLPAGTKYVAFRHYNCSDLNVILLDDIQFTMGGSP 371
             ::   ::: | | :||: : : :||
Db   812 ---AAPQANAKIWIAGQGPTKEDDYVFEAGKKY----HFLMKKMGSGDGTELTISEGG- 863

Qy   372 TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPV 431
          |||||||||||||||||||||||||||||||||||||||||||:|||| |:|||:||
Db   864 ---SDYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVEVKAAGVSPKVCVDYIPDGV 921

Qy   432 QFNPVQN-LTGSAVGQKVTL 450
          ||||||| :|:| |:||:|:
Db   922 ADVTAQKPYTLTVVGKTITV 941

RESULT 6
AF2959
conserved hypothetical protein Atu3276 [imported] - Agrobacterium tumefaciens (strain C
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AF2959
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, C.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF2959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1052 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44092.1; PID:g17741659; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3276
A;Map position: linear chromosome

Query Match   5.6%;   Score 139.5; DB 2;   Length 1052;
Best Local Similarity   20.9%; Pred. No. 0.12;
Matches   123; Conservative   67; Mismatches   159; Indels   239; Gaps   33;

Qy   13 TTTLSESFENGIPASWKTIDADGDGNNWTT-----------TPPPGGTSFAGHNSAICASSAS 64
          |::: ||||||||| ||:|::|||
Db   52 TVTSGEAIIGRHSASEATV--TGDGSKWTTGDLQVGGDFSDPGG--LAGNGT------------ 99

Qy   65 YINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGN------ 116
Db 100 ---------------------LNVTAGGSV---DSTVA-HLGVVAGATGSAIVDGKGS 132

Qy 117 ----DASNF------ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVA 166
Db 133 VWTVDRNSLEVGVSGAGSL---AVTGGGLVDAANIIIGTNTGG---NGSVRVSGADSTVK 186

Qy 167 FRHFGCTDFFWINL------DDVEIKANG---KRADFTETFESSTHGEAPAEWTTIDADGD 218
Db 187 SR-----SD--LNVGLYGNGSMTVEAGGAVKSRDGYVATYGGST------SAVTVTGD 231

Qy 219 GQGWLCLSSGQLGWLTAHGGT-NVVASFSWNGMALNPDNYLI------SKDVTGA-TK 268
Db 232 GSSWAMTGTFFVGY--ASGATGNVTVS---NGGAIRATGVTLGDLAGASGTMTITGAGSK 286

Qy 269 VKXYY----------DHY-------------AVMISKTGT--- 293
Db 287 VTAYVDNGTVNSGSVDVGFQGSGSLSVVNGGSLDAINLYVGNALGSSGAVLVSGVGSHVS 346

Qy 294 -NAGD------FTVVFEETPNGINKGGARFGLSTEADGAKPQSV 330
Db 347 VDGLMVVGNAGNGSVEITGGASLAAPTILIATEAGSTGVLSIGAGSGQTARSAGA----- 401

Qy 331 WIE-RTVDLPAGTKYVAFRHYNCSDLNYILLDDIQ---------FTMGGSPTPDYTY 378
Db 402 -VEARAIAFGAGNGSIVFNH--SETGYTLSADISGAGRVVAEAGVTTLSGNNS------- 451

Qy 379 TVYRDGTKIKEGLTETT-------------FEEDG------VA 402
Db 452 --YSGGTTISAGMLKGTAKSFGSGGIVNNAELVVDGGGTLSNAISGTGSFEKTGDGNLLL 509

Qy 403 TGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLTGSAVGQKVTL 450
Db 510 TGNSTYSGATAVSAG------KLSVNGSLASAVSVGSGATVGGTGTI 550

RESULT 7
H98323
hypothetical protein AGR_L_3085 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: H98323
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: H98323
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1341 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK90114.1; PID:g15160106; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_3085
A;Map position: linear chromosome

Query Match   5.6%;   Score 139.5;   DB 2;   Length 1341;
Best Local Similarity   20.9%;   Pred. No. 0.16;
Matches 123; Conservative 67; Mismatches 159; Indels 239; Gaps 33;

Qy   13 TTTLSESFENGIPASWKTIDADGDGNNWTT------TPPPGGTSFAGHNSAICASSAS 64
Db 341 TVTSGEAIIGRHSASEATV--TGDGSKWTTGDLQVGGDTSDPGG--LAGNGT------ 388

Qy   65 YINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGN------ 116
Db 389 ---------------------LNVTAGGSV---DSTVA-HLGVVAGATGSAIVDGKGS 421

Qy 117 ----DASNF------ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVA 166
Db 422 VWTVDRNSLEVGVSGAGSL---AVTGGGLVDAANIIIGTNTGG---NGSVRVSGADSTVK 475

Qy 167 FRHFGCTDFFWINL------DDVEIKANG---KRADFTETFESSTHGEAPAEWTTIDADGD 218
Db 476 SR-----SD--LNVGLYGNGSMTVEAGGAVKSRDGYVATYGGST------SAVTVTGD 520

Qy 219 GQGWLCLSSGQLGWLTAHGGT-NVVASFSWNGMALNPDNYLI------SKDVTGA-TK 268
Db 521 GSSWAMTGTFFVGY-ASGATGNVTVS---NGGAIRATGVTLGDLAGASGTMTITGAGSK 575

Qy 269 VKXYY----------DHY-------------AVMISKTGT--- 293
Db 576 VTAYVDNGTVNSGSVDVGFQGSGSLSVVNGGSLDAYNLYVGNALGSSGAVLVSGVGSHVS 635

Qy 294 ------NAGD------FTVVFEETPNGINKGGARFGLSTEADGAKPQSV 330
Db 636 VDGLMVVGNAGNGSVEITGGASLAAPTILIATEAGSTGVLSIGAGSGQTARSAGA----- 690

Qy 331 WIE-RTVDLPAGTKYVAFRHYNCSDLNYILLDDIQ---------FTMGGSPTPDYTY 378
Db 691 -VEARAIAFGAGNGSIVFNH--SETGYTLSADISGAGRVVAEAGVTTLSGNNS------- 740

Qy 379 TVYRDGTKIKEGLTETT-------------FEEDG------VA 402
Db 741 --YSGGTTISAGMLKGTAKSFGSGGIVNNAELVVDGGGTLSNAISGTGSFEKTGDGNLLL 798

Qy 403 TGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLTGSAVGQKVTL 450
Db 799 TGNSTYSGATAVSAG------KLSVNGSLASAVSVGSGATVGGTGTI 839

RESULT 8
B75622
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: B75622
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75622
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-691 <WHI>
A;Cross-references: GB:AE001826; NID:g6460827; PIDN:AAF12628.1; PID:g6460924; TIGR:DRB0(
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRB0037
A;Map position: megaplasmid
A;Genome: plasmid
A;Note: plasmid MP1

Query Match   5.3%;   Score 132.5;   DB 2;   Length 691;
Best Local Similarity   21.9%;   Pred. No. 0.21;
Matches 111; Conservative 46; Mismatches 178; Indels 171; Gaps 23;

Qy   4 NPNPNPNPGTTLSESFENGIPASWKTIIDADGDGNNWTTTP----------PPGGTSF 51
Db 230 SPNPLPPTGTTT-----------PGTGTGPVGSGNGPAESTGPAYTGIGSTSGTQPPVVVDA 280

Qy   52 AGHNSAICASSASYINFEGP----------QNPD----NYLVTPELS-LPNGGTLTFWVCAQDA 100
Db 281 SGNQYAYPKADAETVNPDSVTMTTIITNPNPTAATYELVPDLSGLPAGVTVIF----TDA 336

Qy 101 N------YASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQ 153
Db 337 NGNFLPDTDGNGRPEVTAGPNGGTA--------------TYRVVVTYPDTESAAAVAG--- 380

Qy 154 KTVQLPAGTK------YVAFRHFGCTDFFWINLDDVEIKANGKRADFT---- 195
Db 381 -PIRIPVGVDGNRDGIVDATVTYNVLLSNLKFGNTNGTALGVSDVPVTRTVTQPSVTTAV 439

```
Db   1044  TPSTPLADGTVVNATATDPAG-NTGGQG---STTVDAIAPAT-----PTVNLSNGSSLSG- 1094
Db    348  RHYNCSDLNYILLDD------IQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGV 401
Db   1095  ---TABPGSTVILTDGNGNPIAEVTADGS--GNWTYT--PSTPIANGTVVNVAQD-- 1143
Qy    402  ATGNHEYCVEVKYTAGVSPKECVN----------VTVDPVQFNPVQNLTGSAVG 445
             |||                               :|||
Db    493  --IAS----------------TGAVPVKTEKTV------YAVVTIPAGQAPGDYM 523
Qy    300  VVFEETPNGINKGGARFGLSTEADGAKPQ-SVWIERTVDLPAGT----------- 342
Db   1144  ASGNSSPPATVTVDSSAPPAPVINPSNGVVISGTAEAGATVTLTDAGGNPIGQVTADGSG 1203

RESULT 10
C89874
autolysin [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: C89874
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: C89874
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1248 <KUR>
A;Cross-references: GB:BA000018; PID:g13700854; PIDN:BAB42150.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: atl

Query Match          5.2%;  Score 128.5;  DB 2;  Length 1248;
Best Local Similarity  22.3%;  Pred. No. 0.84;
Matches 105;  Conservative 47;  Mismatches 165;  Indels 153;  Gaps  26;

Qy     3  PNPNPNPNPGTT-----LSESFENGIPASWKTIDADGDG---NNWTTTPPP----------- 46
Db    586  PTPTPIPKPSTPTNNKLTVSSLNGV---AQINAKNNGLFTTVYDKTGKPTKEVQKTFA 641
Qy     47  -----GGTSF--AGHNSA---ICASSASYINFEGEQN-PDNYLVTPE--LSLPN 87
             |.|                                         |:|:
Db    642  VTKEASLGGNKFYLVKDYNSPTLIGWVKQGDVIYNNAKSPVNVMQTYTVKPGTKLYSVPW 701
Qy     88  G--------GTLTFWVCAQDANYASEHYAVYASSTGNDASNFAN------------ALLE 127
             :|
Db    702  GTYKQEAGAVSGTGNQTFKATKQQQIDKS---YLFGTVNGKSGWVSKAYLAVPAAPKK 757
Qy    128  EVLTAKT-------VVTAPEAIR------GTRVQGTWYQKVQLPAGTKYVAFRHFG 171
             |:|
Db    758  AVAQPKTAVKAYTVTKPQTTQTVSKIAQVKPNNTGIRASVYEKTAK--NGAKY-ADRTFY 814
Qy    172  CT-----------------DFFWINLDDVEIKANGKRADFTE--TPFSSTHGEAPAE 209
Db    815  VTKERAHGNETYVLLNNTSHNIPLGWFNVKDLNVQNLGKEVKTTQKYTVNKSNNGLSMVP 874
Qy    210  WTTIDADGDGQGWLCLSSGQLWLTAHGGTNVVASFSWN------GMALNPDNYLISKDV 263
Db    875  WGT------KNQVILTGNNI----AQGTFNATKQVSVGKDVYLYGTINNRTGWNAKDL 923
Qy    264  TGATKVK-------YYYAVNDGFPGDHYAVMISKTGT----NAGDFTVVFEETPN 307
Db    924  TAPTAVKFTTSAAKDYNYTYVIKNG-NGYYYVYTPNSDTAKYSLKAFNEQPFAVVKEQVIN 982
Qy    308  G---------INKGGARFGLSTEADGAK----------PQSVWIERTVDLPAGTKY 344
             |:|
Db    983  GQTWYYGKLSNGKLAWIKST--DLAKELIKYNQTGMTLNQVAQIQAGLQY 1030

RESULT 11
AH2515
hypothetical protein alr7304 [imported]  - Nostoc sp. (strain PCC 7120) plasmid pCC7120a
C;Species: Nostoc sp. PCC 7120
```

```
Qy    196  ---ETFESSTH-----------GEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGT 239
Db    440  VFPMDLFNDGAYDGNYALSGSTPIGPVKYYATNPDTDGDG----VLSPAELAALPAE--- 492
Qy    240  NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKGTNAGDFT 299
             |||
Db    493  --IAS----------------TGAVPVKTEKTV------YAVVTIPAGQAPGDYM 523
Qy    300  VVFEETPNGINKGGARFGLSTEADGAKPQ-SVWIERTVDLPAGT----------- 342
Db    524  VT--QTATGSLSGTTKGSFNTDKVTVTSPNGSLLIAKRVTTPGTTTPSLNATANPGDAVSY 581
Qy    343  KVVAFRHYNCSDLNYILLDDIQFTMG--------GSPTPTDYTYTV-------YRD- 383
Db    582  TVTATNNYNTSLYGLVLRDPSSSNNLGSFSSSNVFGFIKPSSLRATVSGVSGATVLYRTSN 641
Qy    384  -GTKIKEGLTE--TTFEEDGVATGNH 406
             ||||:
Db    642  LNTWAAQPTVDANTTWVEVGVDTNNN 667

RESULT 9
A83412
hypothetical protein PA1874 [imported]  - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83412
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83412
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2468 <STO>
A;Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1874

Query Match          5.3%;  Score 131.5;  DB 2;  Length 2468;
Best Local Similarity  22.2%;  Pred. No. 1.3;
Matches 120;  Conservative 47;  Mismatches 184;  Indels 189;  Gaps  30;

Qy     5  PNPNPNPGTT-----------TLSESFENGIPASWKTIDADGDGN----------- 38
Db    754  PVINPSNGTTLSGTAEPGSSVTLITDG--NGNPIG--QVTADGSGNWSFTPSTPLADGTVV 809
Qy    39  NWTTTPPPGGTSFAGHN-----------------SAICASSASYINFEGEQNP---- 74
             ||||
Db    810  NATATDPAGNTSGQGSTTVDGVAPTTPTVNLSNGSSLSGTAEPGSTVILTDGNGNPIAEV 869
Qy    75  -----DNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEV 129
Db    870  TADGSGNWTYTPSTPIANGTVVN--VVAQDA----AGN--SSPGASVTVDSQ 913
Qy    130  LTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANG 189
Db    914  APAAPVVNPS--NGTTLSGT------AEPGAT----------VTLTD--GNG 945
Qy    190  KRADFTEFESSTHGEAPAEWTTIDADGDGQGW---LCLSSGQLWLTAH--GGTNV 241
Db    946  NPIG--------QVTADGSG-NWSFTPGTPLANGTVVNATASDPTGNTSA 986
Qy    242  VASFSWNGMA----LNPDNYLVTGATKVKYYYAVNDGFPGDHYAVMISKTG---- 292
Db    987  PASTTVDSVAPAAPVVNPSN---GAEISGTAEPGATVTLTDGSGSGNWSF 1043
Qy    293  ---TNAGDFTVV--FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAF 347
             |||:
```

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AH2515
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
,N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2515
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-4936 <KUR>
A;Cross-references: GB:BA000020; PIDN:BAB78388.1; PID:g17135842; GSPDB:GN00180
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr7304
A;Genome: plasmid

Query Match   5.2%;   Score 128;   DB 2;   Length 4936;
Best Local Similarity   20.4%;   Pred. No. 5.5;
Matches   108;   Conservative   60;   Mismatches   171;   Indels   190;   Gaps   27;

Qy   1 GTPNPNPNPNPDGT-------TTLSESEFENGIPASWKTIDADGDGNNWTT-----TPPPGG 48
                                 : :::: || || : : |
Db   3161 GLSNFN--TPGTYQLQVDATTVQDNAGN----------SGDAARTTFTTIAAPPTPGV 3206

Qy   49 TSFAGHNSAICASSASYINFRGPQNPDNYLV------TPELSLP-NGG------TLTF 93
           | :: :::| |||:
Db   3207 T----ITQSGGSTAVIEG-GNTDSYTLVLRTQPTADVTVTLNTGSQITTDKTTLTF 3257

Qy   94 ----WVCAQ--------DANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEA 141
                            |:| | | ||||||
Db   3258 TSANWNTPQTITVNAVNDTITEGNHTSTISHSISSTDTNYSNVTLPDI--AVSITDNDAE 3315

Qy   142 IRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGKRA--DFTETFE 199
           || | ::| || | |||||||||||||||| ::   : | | |
Db   3316 IRGMK-------------WNDIDGDGVKDTGEPGLQGWTIYLD 3345

Qy   200 SSTHGEAP--AEWTTIDADGDGQGWLCLSSG--------QLGWLTAHGGTNV------ 241
           :::| || |||||||||||| ||||: |||||
Db   3346 SNTNGQLDNGEISTTTDANGNVQ-FTNLRPGVYTVAEVQQPGWKQTFPGTNITTTNADIP 3404

Qy   242 --VASFSWNGMALN--PDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKT 291
             ||| | || ||||| || ||| || | | ||||
Db   3405 LAIPSLDMISPGDSNGIQLNFSAANYIVKEDGTAITEV--WVTRTGNTSSAVSATLSFT 3461

Qy   292 -GTNAG-----------DFTVVFEE------------TPNGIN---- 310
            || | || :: | ||| |
Db   3462 DGTATGCGCGASSVNNDFNNVPFTIAFAENETSKLLSVQNALLANPNAIKIRNDSKVEGN 3521

Qy   311 ------KGGARFGLSTEA-----DGAKPQSVWIERTVDLPAG--TKYVAFRHYNCS 353
                 ||||||||| || | | | | | |
Db   3522 EYFTIKLTNPTGGAVIGNQSIATVTIIDDEAPSDITVTPPLETPSTTITSAVDSQAIYLI 3581

Qy   354 DLNYILLDDIQFTMGGSPTPDYTYVYRDGTKIKEGLTETFEEDGVA 402
           | : | || || |||||||||:||||:| | : | | |
Db   3582 NLNNFWADSRFANIKGN-----DFTSVIIDTGIDLNFPFGADTDNNGIA 3626

RESULT 12
S10789
amylase A-180 - alkaliphilic eubacterium 163-26
C;Species: alkaliphilic eubacterium 163-26
C;Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 15-Oct-1999
A;Accession: S10789
R;Candussio, A.; Schmid, G.; Boeck, A.
Eur. J. Biochem. 191, 177-185, 1990
A;Title: Biochemical and genetic analysis of a maltopentaose-producing amylase from an a
A;Reference number: S10789; MUID:90336627; PMID:1696201
A;Accession: S10789
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1684 <CAN>

A;Cross-references: EMBL:X53373; NID:g48305; PIDN:CAA37453.1; PID:g48306

Query Match   5.1%;   Score 127.5;   DB 2;   Length 1684;
Best Local Similarity   20.9%;   Pred. No. 1.5;
Matches   102;   Conservative   53;   Mismatches   184;   Indels   149;   Gaps   27;

Qy   23 GIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA-SSASYINFEGPQNPDNYLVTP 81
           ||: ||:
Db   -----------HNDIMNKDNEAAWANWWGSD------- 241

Qy   204 GLPRDWTPNQA--QGQNWHT 241
           : | || | | | |
Db   82 ELSLPNGGTLTFWVCAQD-ANYASEHYAVYASSTGNDASNFANALLEEVLTAKTV-VTAP 139

Qy   242 ------------WIRADETAGY----DNCGGSEQTMCIGFLPDIKTEVTTGVDLP 280
                      :: : | | || : | | : | | |
Db   140 EAIRGT-RVQGTWYQKTVQLPAGTKYVAFRHFGCTDFF--WINLDDVEIKANGKRADFTE 196

Qy   281 PILRNKWNDQASGYEDWF-VPAAEPYRQDLNIAPKDYLIKWITSWVEBFGIDGFRVDTAK 339
           ::| | | || | : | :| :| | | | : | | |
Db   197 TFESSTHGEAPAE--------WTTIDADGDGQGWLCLSSGQLGWLTA-----HG------ 237

Qy   340 HVEIERWAELKNEAEVALQTWRENNPDKPGANW------DDNFWMTAEVFGHGLGKSEVFD 394
                | | | || | | |
Db   238 -GTNVVASFSW-----------------NGMALNPDNYLISKDVTGATKVKYYYAVNDGFP 280

Qy   395 FGFDSVINFEFQNANFNNLEGLFSRYANSINTDPDFNMLSYVSSHDTKL--YSRDD---- 448
              || | :: | | | | | | ||: | :| | |
Db   281 GDHYAVMISKTGTN-----AGDFTVVF-EETPNGINKGGARFGLSTEADGAKPQSVW--IE 333

Qy   449 -----LIQAGTALLLLPGGVQVFYGDETARPLGDGG-----SDPEQGTRSSMNWANIN 496
                |: | ||| | | | | | | || |
Db   334 RTVDLPAGTKYVAFR--------HYNCSDLNYI--------LLDDIQFTMGGSPTPT 374

Qy   497 QNV-LSHWQKLGQFRNNHIAIGAGAHQKLSDSPYTFARTYESDDIVDEVVATGAQGTTA 555
             | | | | | | | | | : | |
Db   375 DYTYTVYRDGTKIKEGLT--ETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPVQ 432

Qy   556 VTVEGVFEDGTVVRDAYTGDETTVIK-GTAT----------FTAGTQGIILIENTARPVT 604
                | | | : | | :: | | |
Db   433 FNPVQNLT 440

Db   605 NLPIVSAT 612

RESULT 13
B75489
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: B75489
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
,M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-713 <WHI>
A;Cross-references: GB:AE001925; GB:AE000513; NID:g6458883; PIDN:AAF10267.1; PID:g64583
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0685
A;Map position: 1

Query Match   5.1%;   Score 127;   DB 2;   Length 713;
Best Local Similarity   21.6%;   Pred. No. 0.51;
Matches   129;   Conservative   47;   Mismatches   214;   Indels   208;   Gaps   29;

Qy   2 TPNPNP-NPNPGTTLSE----SFENGIPASWKTIDADGDGNNWTTTPP----PGGT---- 49
             |:|| ||||| | : | | |||
Db   86 TPESTPTNPNPPSEIETPVITVNPVPSFTIT---PNDGSKDTQPDYTNPGQTREVK 142

```
Qy   50 ------SFAGHNSAICASSASYINFE-----GPQNPDNYLVTPE------LSLPNGGTL--   91
                       :| | :|| :  |||
Db  143 PCDKNVTF------AYTLTNTGNVNGESYTILTNTPDTGAVKTPENIRFYLDSNNNGQLDQ  197

Qy   92 -----TFWVCAQDANYASEHYAVY------------ASSTGNDASNFAN--ALLE        127
             :| | : ||
Db  198 SEIAAGATTTITDVAINQTVKFFQVVDVPCTATSTDKFGGDPTGTRNDNPNFSNDPTLPR   257

Qy  128 EVLTAKTVV------------TAPEAIRGTRVQGTWYQKTV-QLPA                160
         :  |||
Db  258 DANNSNTVTINRKDGVVIGPKADPDGNGNPVTPAYNSPEGINIVPTASDTQVATVTTLPT   317

Qy  161 GTKYVAFRHFGCTDFFWINLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQ   220
          :
Db  318 SGVTVTFTN------TIQNTGNR---TDTFELTQTNTFPAGTTVVFKDANGN          360

Qy  221 GWLCLSSGQLGWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFP   280
           :
Db  361 A------LPTVNGKPVVQNVPENGTV---DIQVIVTLPAGVTPTQL-----SGQP       401

Qy  281 GDHYAVMISKTGTN---AGDFT--VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERT   335
           :
Db  402 ---AVTVTTTSQNDPTKSDTTKDIIEVKVP------GIAFGDFTPGLGGDPTPVGTPPT    451

Qy  336 ------VDLPAG---TKYVAFRHYNCSDLNYILLDDIQPTMGG---------SPT       372
              :||
Db  452 GVPGNPGTPLTPGNPQTCTAPIRTYLPMEIANLGSQODDAFVVSGTAPVTVLNPDGTVNPT  511

Qy  373 PTDYTYTVYRD---GTKIKBGLTETTFEBDGVATGNHEYCVEVKY------           414
            ||| |
Db  512 PVIVPVVYRDVNGDGKLDAGDTEL---QGGNTGTIKPGEEIKLIAVVDVPCAAAQQTI     567

Qy  415 ------TAGVSPKE------CVNVTVDPVQFNPVQNLTGSAVGQKVT               449
               ||||:|
Db  568 TLNQEAKSPTTGVSQKDPNDTITVGGNGGKPIVKTVDKATANPGDTLTYTIIGKNTS      625
```

RESULT 14
A86685
prophage pil protein 46, tail component [imported] - Lactococcus lactis subsp. lactis (s
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: A86685
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86685; MUID:21235186; PMID:11337471
A;Accession: A86685
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1441 <STO>
A;Cross-references: GB:AE005176; PID:g12723361; PIDN:AAK04579.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: pil46

```
Query Match           5.1%;  Score 126;  DB 2;  Length 1441;
Best Local Similarity  22.3%;  Pred. No. 1.5;
Matches    98;  Conservative   49;  Mismatches  133;  Indels  160;  Gaps  26;

Qy   30 TIDADGDGNNWTTTPPGGTSFAGHNSAICASSASYINPEGPQNP----DNYLVTPELSL    85
                                :  |  |||||
Db  602 TISSSG-----TVTPTTGWTS------QVPTLVKEQYL------                 628

Qy   86 PNGGTLTFWVCAQDANYASEHYAV-YASSTGN------DASNFANALLEEVLTAKTV     135
            |||
Db  629 ---WTKTLWTYTD--NTSETGYSVSYIAKDGNNGHDGFPGKDGVGISNTIIEYV---     677

Qy  136 VTAPEAIRGT-RVQGTWYQKTVQLPAGTKYVAFRHF-----GCTDFFWINL---        180
            ::      |||        :|
Db  678 ----GAVSGTSKPTGGWSTTIPTVPAG-QYLWTRTTWQYTDGTSEQGYINALMGLTGASG   732

Qy  181 -DDVEIKANGKRADFTE-TFESSTHG-EAPAE------WTTIDADG                217
```

```
Db  733 RDGIAGK-DGKGIKATAITYQASTNGTTAPTGTWSTSVPSVAKGSFLWTRTIWTYTD--    788
          | : ||
Qy  218 DGQGWLCLSSGQLGWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVN-   276
            : |: : |
Db  789 ------NTTETGYAVAYMGYN---GNNGTNGIAGKDGTGIKTTTITYAVGT           830

Qy  277 ------DGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEA          322
             : |||
Db  831 SGTTAPTGGWNSQVPNVPAGQY-LWTKT--VWDYTDKTSETGYSVSKFGEK-GDKGDQ     884

Qy  323 DGAKPQSVWIERTVDLPAG----TKYVAFRHYNCSD-LNYILLDD------IQFTMG     368
           :
Db  885 GVQGIQGVDGRQGIPGPKGADGKTQYTHIAYANSADGKTNFSTSDSNRTYIGMYVDFNIN   944

Qy  369 GSPTPTDYTYTVYR--DGTK                                          386
         |||:|:|: |
Db  945 DSTTPSDYSWTLVKGADGTQ                                          964
```

RESULT 15
C86822
hypothetical protein yqbK [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: C86822
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrl
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: C86822
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1649 <STO>
A;Cross-references: GB:AE005176; PID:g12724583; PIDN:AAK05677.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yqbK

```
Query Match           5.1%;  Score 126;  DB 2;  Length 1649;
Best Local Similarity  20.4%;  Pred. No. 1.8;
Matches   110;  Conservative   56;  Mismatches  189;  Indels  184;  Gaps  28;

Qy   46 PGGTSFAGHNSAICASSASYINFEG-PQN----------PDNY--------LVTPELSLPN  87
          | :  |: :
Db   65 PSGTTFNG------NSVSQNRTPLPQNSVEDAIMPNSFGWQQTIANLLPTNQVTAVQ      117

Qy   88 GGTLTFWVCAQDANYASEHYAVYASSTGNDASNF------ANALLEEVLTAKTVVTAPEA- 141
          |  |  : :
Db  118 NGTIIGSYPTTDAGVQQALYAMY-SATNASTSDFGLYGANTTLSNSL-LKAVVSNPSAT    175

Qy  142 ------IRGTRVQGTWYQKTVQL-------------------PAGTKYVAFR---HFGCTDFFWINLDDV   183
           :|
Db  176 NMTFSTLKGHAKSLTWISNPADLLTSSNSSQPSGTNYTGTLPSNVYFGVPTIF---RNV    231

Qy  184 EIKANGKRADFTETFESSTHGEAPAEWTT----------IDADGDGQGW             222
          : |
Db  232 TVAASGDNIYAQGNAFATTNG------SWITGAPNIYGGTDNSDISGNTNLYIGATGSIAGW 287

Qy  223 LCLSSGQLGWLTAHGGTNV------VASFSWNGMALNPDNYLISKDVTG-ATKV         269
          :
Db  288 -NIYGGNASAATISGNTHVTIAQSSTINSVTGGSASGTTISGNTNL--DISGAIASQI     343

Qy  270 KYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNG-----INKGGARFG----LS   319
                                                  : |
Db  344 TNIYGAGIG--------TSNSPVNVGNVTTYVNSTNGGARYQLYQGGTVYGNISGSIY     394

Qy  320 TEADGAKPQSVWIERTVDL------PA----------GTKYVAFRHYNCSDLNYIL       359
                              ::      ::
Db  395 NTLSGA--GGWTGATSNINGAGGPASTNGGSFOGNIGTSGAGNVISNSYNTSSFT---    448

Qy  360 LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAG-   417
```

```
Db   449 TGQALFTGGNAGTGASYAQA----------------TNSTTAAQGILYANITNYIKSAFTTGTA 496

Qy   418 ------------VSPKE----------CVNVTVDPVQFNPVQNLTGSAVGQKVT 449
                     :|| :          :|  :  :  :  | :  |||:|:
Db   497 GAVYGIVGGNGHDSLKISPSQWGLGSQTGLDSAVGVT-DAKAYGQIPSTTVVSNAQKIT 554
```

Search completed: May 18, 2004, 11:47:50
Job time : 12.906 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May 18, 2004, 11:35:14 ; Search time 8.3342 Seconds
                                       (without alignments)
                              2848.981 Million cell updates/sec

Title:          US-08-570-311-20
Perfect  score: 2480
Sequence:       1 GTPNPNPNPNPGTTTLSESF.........QNLTGSAVGQKVTLKWDAPN 456

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:        141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

         %
Result          Query
No.    Score   Match  Length DB  ID              Description
-----  -----   -----  ------ --  --              -----------
  1    2480    100.0   2628  1   HGA2_PORGI      Q51845 porphyromon
  2    2441     98.4   2164  1   HGA1_PORGI      P59915 porphyromon
  3    1453     58.6    989  1   PRTH_PORGI      P46071 porphyromon
  4     266     10.7    991  1   CPG1_PORGI      P28784 porphyromon
  5    131.5     5.3   1256  1   ATL_STAAU       P52081 staphylococ
  6     126      5.1    721  1   OGP_MOUSE       Q62010 mus musculu
  7    121.5     4.9    465  1   SLAP_LACBR      Q05044 lactobacill
  8     121      4.9   1090  1   GUXB_CELFI      P50899 cellulomona
  9    118.5     4.8   1385  1   C5AA_BACUD      Q45760 bacillus th
 10     114      4.6    872  1   GUXA_CELFI      P50401 cellulomona
 11     113      4.6   2660  1   YEEJ_ECO57      Q8X8V7 escherichia
 12    112.5     4.5   1034  1   BGAL_BACME      O52847 bacillus me
 13    112.5     4.5   1260  1   ALS1_CANAL      P46590 candida alb
 14    110.5     4.5   1122  1   ADP1_MYCGA      Q49379 mycoplasma
 15    109.5     4.4    331  1   PME_ASPAC       Q12535 aspergillus
 16    109.5     4.4    607  1   YK28_ARCFU      O28251 archaeoglob
 17    108.5     4.4   1953  1   BIGA_SALTY      P25927 salmonella
 18     108      4.4    699  1   CHI1_BACCI      P20533 bacillus ci
 19     108      4.4    987  1   YD94_METJA      Q58789 methanococc
 20     108      4.4   1289  1   C5AB_BACUD      Q45753 bacillus th
 21    107.5     4.3    551  1   AMYB_THETU      P19584 thermoanaer
 22    107.5     4.3   1200  1   HYAL_STRPU      O76536 strongyloce
 23     107      4.3   1656  1   OMPB_RICJA      O06653 r outer mem
 24     107      4.3   2249  1   OMPA_RICRI      P15921 rickettsia
 25    106.5     4.3   1122  1   ADP2_MYCGA      Q9rem8 mycoplasma
 26    106.5     4.3   2124  1   PGCA_RAT        P07897 rattus norv
 27    106.5     4.3   2358  1   YEEJ_ECOLI      P76347 escherichia
 28     106      4.3   1045  1   PRTS_SERMA      P09489 serratia ma
 29     106      4.3   1300  1   120K_RICRI      P14914 rickettsia
 30     106      4.3   1654  1   OMPB_RICRI      Q53047 r outer mem
 31    105.5     4.3   1220  1   C5AC_BACTU      P56955 bacillus th
 32    105.5     4.3   1534  1   YFAS_ECO57      Q8X835 escherichia
 33     105      4.2    827  1   XANP_XANS2      Q60106 xanthomonas
 34      105      4.2  1045  1   GUNB_CELFI      P26225 cellulomona
 35      105      4.2  2812  1   ZAN_HUMAN       Q9Y493 homo sapien
 36     103.5    4.2   837  1   XYNZ_CLOTM      P10478 clostridium
 37     103.5    4.2  1655  1   OMPB_RICCN      Q9kka3 r outer mem
 38     103.5    4.2  1861  1   APU_THETU       P38536 t amylopull
 39      103     4.2  1157  1   C8AA_BACUK      Q45704 bacillus th
 40     102.5    4.1   282  1   PRTA_ASPNG      P24665 aspergillus
 41     102.5    4.1  1045  1   PRTT_SERMA      P29805 serratia ma
 42     102      4.1   435  1   AM3D_ORYSA      P27933 oryza sativ
 43     102      4.1   941  1   GUN_BAC56       P19424 bacillus sp
 44     101.5    4.1  1462  1   PTP6_DROME      P16620 drosophila
 45     101.5    4.1  2132  1   PGCA_MOUSE      Q61282 mus musculu

                              ALIGNMENTS

RESULT 1
HGA2_PORGI
ID   HGA2_PORGI      STANDARD;      PRT;   2628 AA.
AC   Q51845;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97047672; PubMed=8926061;
RA   Han N., Whitlock J., Progulske-Fox A.;
RT   "The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381
RT   contains four large, contiguous, direct repeats.";
RL   Infect. Immun. 64:4000-4007(1996).
CC   -!- FUNCTION: Agglutinates erythrocytes.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   -----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U41807; AAB17128.1; -.
DR   PIR; T28651; T28651.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25   2628       HEMAGGLUTININ A.
FT   DOMAIN      25    539       PEPTIDASE C25-LIKE 1.
FT   DOMAIN     540    995       PEPTIDASE C25-LIKE 2.
FT   DOMAIN     996   1451       PEPTIDASE C25-LIKE 3.
FT   DOMAIN    1452   1907       PEPTIDASE C25-LIKE 4.
FT   DOMAIN    2074   2628       PEPTIDASE C25-LIKE 5.
SQ   SEQUENCE   2628 AA;  283324 MW;  61C4DB32540C99DA CRC64;

Query Match             100.0%;  Score 2480;  DB 1;  Length 2628;
Best Local Similarity   100.0%;  Pred. No. 1.6e-163;
Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1  GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFPAGHNSAICA  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   950  GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFPAGHNSAICA 1009

Qy    61  SSASYINFEGPQNPDNYLVTPELSLPNGGTLFWVCAQDANYASEHYAVVASSTGNDASN  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1010  SSASYINFEGPQNPDNYLVTPELSLPNGGTLFWVCAQDANYASEHYAVVASSTGNDASN 1069
```

```
Best Local Similarity 98.9%; Pred. No. 6.2e-161;
Matches 449; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy    3 PNPNPNPGTTLSESPENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASS   62
          ||||||||||||||||||||||||||||||||||||||||| |||||||||| |
Db  496 PNGTPNPNPGTTLSESPENGIPASWKTIDADGDGNNWTTTPPGGSSFAGHNSAICVSS  555

Qy   63 ASYINFEGPQNPDNYLVTPELSLPNGGLTFWVCAQDANYASEHYAVVASSTGNDASNFA  122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  556 ASYINFEGPQNPDNYLVTPELSLPNGGLTFWVCAQDANYASEHYAVVASSTGNDASNFA  615

Qy  123 NALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDD  182
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  616 NALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDD  675

Qy  183 VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVV  242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  676 VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVV  735

Qy  243 ASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVF  302
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  736 ASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVF  795

Qy  303 EETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDD  362
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  796 EETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDD  855

Qy  363 IQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKE  422
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  856 IQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKE  915

Qy  423 CVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  456
          ||||||||||||||||||||||||||||||||||
Db  916 CVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  949


RESULT 3
PRTH_PORGI
ID   PRTH_PORGI     STANDARD;      PRT;    989 AA.
AC   P46071;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protease prtH (EC 3.4.22.-).
GN   PRTH.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=95012612; PubMed=7927685;
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RT   "Cloning and characterization of a new protease gene (prtH) from
RT   Porphyromonas gingivalis.";
RL   Infect. Immun. 62:4279-4286(1994).
RN   [2]
RP   ERRATUM.
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RL   Infect. Immun. 62:5707-5707(1994).
CC   -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
CC       P.GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
CC       IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
CC       AND IS A VIRULENCE FACTOR.
CC   -!- SUBCELLULAR LOCATION: In membrane vesicles.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1070 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 1129

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1130 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 1189

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1190 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 1249

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1250 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1309

Qy  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1310 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1369

Qy  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  456
          ||||||||||||||||||||||||||||||||||||
Db 1370 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 1405


RESULT 2
HGA1_PORGI
ID   HGA1_PORGI     STANDARD;      PRT;   2164 AA.
AC   P59915;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA OR PG1837.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=22829867; PubMed=12949112;
RA   Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA   Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
RA   Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
RA   Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA   Dewhirst F.E., Fraser C.M.;
RT   "Complete genome sequence of the oral pathogenic bacterium
RT   Porphyromonas gingivalis strain W83.";
RL   J. Bacteriol. 185:5551-5601(2003).
CC   -!- FUNCTION: Agglutinates erythrocytes (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; AE017178; AAQ66831.1; ALT_INIT.
DR   TIGR; PG1837; -.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
KW   Complete proteome.
FT   SIGNAL        1     25                POTENTIAL.
FT   CHAIN        26   2164       HEMAGGLUTININ A.
FT   DOMAIN       26    539       PEPTIDASE C25-LIKE 1.
FT   DOMAIN      540    991       PEPTIDASE C25-LIKE 2.
FT   DOMAIN      992   1443       PEPTIDASE C25-LIKE 3.
SQ   SEQUENCE   2164 AA;  233387 MW;  6DFAB22832586C63 CRC64;

Query Match  98.4%;  Score 2441;  DB 1;  Length 2164;
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L27483; AAA51298.1; -.
DR   HSSP; P23382; 1FMT.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR002376; formyl_transf.
DR   Pfam; PF00551; formyl_transf; 1.
KW   Hydrolase; Thiol protease; Repeat; Virulence.
FT   REPEAT        270    323
FT   REPEAT        528    581
SQ   SEQUENCE   989 AA;  110238 MW;  FA85FE8A3AC8944C CRC64;

Query Match             58.6%;   Score 1453;  DB 1;  Length 989;
Best Local Similarity   84.9%;   Pred. No. 5.7e-93;
Matches  276;  Conservative  11;  Mismatches  30;  Indels  8;  Gaps  3;

Qy    136  VTAPEAIRGTRVQGTWVQKTVQLPAGTKYVAFRHFG----CTDFFWINLDDVEIKANGKR  191
                ||||||| |||||||||||||||||  :  |  :|   |  ||
Db     34  VRSPEAIRG-RIQGTWRQKTVDLPAGTEICCFPSLPKAPICSTSTLMRLI---RSKTNAKR   89

Qy    192  ADFTETPESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWNGMA  251
                ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     90  ADFTETPESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVSSFSWNGMA  149

Qy    252  LNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEETPNGINK  311
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    150  LNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEETPNGINK  209

Qy    312  GGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP  371
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    210  GGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP  269

Qy    372  TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPV  431
                |||||||||||||||||||||||||||||||||||||||||||||||::||||:
Db    270  TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTVNST  329

Qy    432  QFNPVQNLTGSAVGQKVTLKWDAPN  456
                |||||:||  ||:::|||||:||:
Db    330  QFNPVKNLKAQPDGGDVVLKWEAPS  354
```

```
RL   J. Biol. Chem. 267:18896-18901(1992).
CC   -!- FUNCTION: Thiol protease which is believed to participate in
CC       intracellular degradation and turnover of proteins. Its
CC       proteolytic activity is a major factor in both periodontal tissue
CC       destruction and in bacterial host defense mechanisms. Activates
CC       complement C3 and C5.
CC   -!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
CC       immunoglobulins, with a preference for Arg in P1, and hydrophobic
CC       residues in P2 and P3.
CC   -!- ENZYME REGULATION: Requires cysteine for activation and Ca(2+)
CC       and/or Mg(2+) for stabilization. It is stimulated by glycine-
CC       containing dipeptides. It is resistant to inhibition by proteinase
CC       inhibitors in human plasma.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; D26470; BAA05484.1; -.
DR   PIR; I40229; I40229.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
KW   Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
FT   SIGNAL          1     24
FT   PROPEP         25    227
FT   CHAIN         228    991       GINGIPAIN R1.
FT   ACT_SITE      438    438       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE      471    471       NUCLEOPHILE (BY SIMILARITY).
FT   CONFLICT      264    265       RT -> TK (IN REF. 2).
SQ   SEQUENCE   991 AA;  108782 MW;  03EE3F43CBBE2544 CRC64;

Query Match             10.7%;   Score 266;  DB 1;  Length 991;
Best Local Similarity   25.9%;   Pred. No. 8.6e-11;
Matches  114;  Conservative  50;  Mismatches  164;  Indels  112;  Gaps  19;

Qy     30  TIDADGDGNWWTTTPPPGGTSFAGHNSAICASSASYINFEGPQNPDNVLVTPELSLPNGG   89
                :::  ::                              :::|
Db    595  SVNVSCDYNGAIATISANGKMF---GSAVVENGTATINLTG----------LTNES  637

Qy     90  TLTFWVCAQDANYASEHYAVYASSTG--NDASNFANALLEEVLTAKTVVTAPEAIRGTRV  147
                |||| ||       : :       :::         :: ||:  :
Db    638  TLTLTV---VGYNKETVIKTINTNGBENPYQPVSN------LTATT----QGQKV  679

Qy    148  QGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV-EIKANGKRADFTETFESSTHGEA  206
                |||:||| |  :|  |  ||:||  :|  : :||  :|:
Db    680  TLKWDAPSTKTNATTN-TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWNDGS-  737

Qy    207  PAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWN----GMALNPDNYLISK  261
                :|||                     ::  |        :|
Db    738  -GYQILLDADHDQYGQVIPSDTHTLWPNCSVPANLFAPFEYTVPENADPSCSPTNMIM--  794

Qy    262  DVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTE  321
                |||:          ::                 ::   |
Db    795  DGTASVNI----------PAGTY--------DFAI----------------  811

Qy    322  ADGAKPQS---VWI------ERTVDLPAGTKVVAFRHYNCSDLNYILLDDIQFTMGGSP  371
                |: | :|  :|        : :|      ||   :|  :||||||||||||||:|
Db    812  ---AAPQANAKIWIAGQGPTKEDDYVFEAGKKY----HFLMKKWGSGDGTELTISEGG-  863

Qy    372  TPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPV  431
                |||||||||||||||||||||||:|  || ||::|||||||||:||| |||:   |
Db    864  --SDYTYTVYRDGTKIKEGLTETTYRDAGMSAQSHEYCVEVKXAAGVSPKVCVDYIPDGV  921
```

```
RESULT 4
CPG1_PORGI
ID   CPG1_PORGI     STANDARD;      PRT;   991 AA.
AC   P28784; Q45168;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
DE   (RGP-1).
GN   RGPA OR RGP1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 228-290 AND 517-541.
RC   STRAIN=381;
RX   MEDLINE=95168884; PubMed=7864651;
RA   Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
RA   Ikehara Y.;
RT   "Structural characterization of arginingpain, a novel
RT   arginine-specific cysteine proteinase as a major periodontal
RT   pathogenic factor from Porphyromonas gingivalis.";
RL   Arch. Biochem. Biophys. 316:917-925(1995).
RN   [2]
RP   SEQUENCE OF 228-270.
RC   STRAIN=HG66;
RX   MEDLINE=92406812; PubMed=1527017;
RA   Chen Z., Potempa J., Polanowski A., Wikstrom M., Travis J.;
RT   "Purification and characterization of a 50-kDa cysteine proteinase
RT   (gingipain) from Porphyromonas gingivalis.";
```

```
Qy    432 QFNPVQN-LTGSAVGQKVTL 450
              |:|||.|| :|:||
Db    922 ADVTAQKPYTLTVVGKTITV 941


RESULT 5
ATL_STAAU
ID   ATL_STAAU      STANDARD;      PRT;    1256 AA.
AC   P52081;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Bifunctional autolysin precursor [Includes: N-acetylmuramoyl-L-alanine
DE   amidase (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
DE   acetylglucosaminidase (EC 3.2.1.96)].
GN   ATL.
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
RC   STRAIN=RN450;
RX   MEDLINE=95116542; PubMed=7816834;
RA   Oshida T., Sugai M., Komatsuzawa H., Hong Y.-M., Suginaka H.,
RA   Tomasz A.;
RT   "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
RT   alanine amidase domain and an endo-beta-N-acetylglucosaminidase
RT   domain: cloning, sequence analysis, and characterization.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 8325-4;
RA   Foster S.J.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOBIOSYL UNIT IN
CC       HIGH-MANNOSE GLYCOPEPTIDES AND GLYCOPROTEINS CONTAINING THE
CC       -[(MAN)5(GLCNAC)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
CC       RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
CC       OLIGOSACCHARIDE IS RELEASED INTACT.
CC   -!- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
CC       residues and L-amino acid residues in certain bacterial cell-wall
CC       glycopeptides.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC       unit in high-mannose glycopeptides and glycoproteins containing
CC       the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC       residue remains attached to the protein; the rest of the
CC       oligosaccharide is released intact.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
CC       EXTRACELLULAR LYTIC ENZYMES.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
CC       ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
CC       GLYCOSYL HYDROLASES.
```

```
DR   EMBL; D17366; BAA04185.1; -.
DR   EMBL; L41499; AAA99982.1; -.
DR   InterPro; IPR002502; Amidase_2.
DR   InterPro; IPR002901; Amidase_4.
DR   Pfam; PF01510; Amidase_2; 1.
DR   Pfam; PF01832; Amidase_4; 1.
DR   SMART; SM00644; Ami_2; 1.
DR   SMART; SM00047; LY22; 1.
KW   Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30   1256       BIFUNCTIONAL AUTOLYSIN.
FT   DOMAIN      199    775       N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
FT   DOMAIN      776   1256       ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
FT   REPEAT      425    589       1.
FT   REPEAT      596    758       2.
FT   REPEAT      770    932       3.
SQ   SEQUENCE   1256 AA;  137384 MW;  2BB76CAA292FDD20 CRC64;

Query Match             5.3%;  Score 131.5;  DB 1;  Length 1256;
Best Local Similarity  22.3%;  Pred. No. 0.24;
Matches 105;  Conservative 47;  Mismatches 165;  Indels 153;  Gaps 26;

Qy    3 PNPNPNENPGTTT---LSESFENGIPASWKTIDADGDG---NNWTTTPPP----------- 46
Db  594 PTPTPTPKPSTPTTNNKLTVSSLNGV---AQINAKNNGLFTTVYDKTGKPTKEVQKTFA 649

Qy   47 ------GGTSF--AGHNSA----ICASSASYINFEGPQN-PDNVLVTPE--LSLPN 87
Db  650 VTKEASLGGNKFYLVKDYNSPTLIGWVKQGDVIYNNAKSPVNVMQTYTVKPGTKLYSVPW 709

Qy   88 G-------------GTLTFWVCAQDANYASEHYAVAYASSTGNDASNFAN-----------ALLE 127
Db  710 GTYKQEAGAVSGTGNQTFKATKQQQIDKS----IYLFGTVNGKSGWVSKAYLAVPAAPKK 765

Qy  128 EVLTAKT----VVTAPEAIR----------------GTRVQGTWYQKTVQLPAGTKYVAFRHFG 171
Db  766 AVAQPKTAVKAYTVTKPQTTQTVSKIAQVKPNNTGIRASVYEKTAK--NGAKY-ADRTFY 822

Qy  172 CT--------------DFFWINLDDVEIKANGKRADFTE--TFESSTHGEAPAE 209
Db  823 VTKERAHGNETYVLLNNTSHNIPLGWFNVKDLNVQNLGKEVKTTQKYTVNKSNNGLSMVP 882

Qy  210 WTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWN-------GMALNPDNYLISKDV 263
Db  883 WGT-------KNQVILTGNNI---AQGTFNATKQVSVGKDVYLYGTINNRTGWVNAKDL 931

Qy  264 TGATKVK----------YYYAVNDGFPGDHYAVMISKTGT------NAGDFTVVFEETPN 307
Db  932 TAPTAVKPTTSAAKDYNYTVIKNG-NGYYVVTPNSDTAKYSLKAFNEQPFAVVKEQVIN 990

Qy  308 G------INKGGARFGLSTEADGAK-----------PQSVWIERTVDLPAGTKY 344
Db  991 GQTWYYGKLSNGKLAWIKST--DLAKELIKYNQTGMTLNQVAQIQAGLQY 1038


RESULT 6
OGP_MOUSE
ID   OGP_MOUSE      STANDARD;      PRT;    721 AA.
AC   Q62010;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Oviduct-specific glycoprotein precursor (Oviductal glycoprotein)
DE   (Oviductin) (Estrogen-dependent oviduct protein).
GN   OVGP1 OR OGP OR CHIT5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR; TISSUE=Oviduct;
RX   MEDLINE=96115001; PubMed=7492680;
RA   Sendai Y., Komiya H., Suzuki K., Onuma T., Kikuchi M., Hoshi H.,
RA   Araki Y.;
RT   "Molecular cloning and characterization of a mouse oviduct-specific
RT   glycoprotein.";
RL   Biol. Reprod. 53:285-294(1995).
CC   -!- FUNCTION: Binds to oocyte zona pellucida in vivo. May play a role
CC       in the fertilization process and/or early embryonic development.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- TISSUE SPECIFICITY: Epithelial cells of the oviduct.
CC   -!- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.
```

```
CC   -------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------
DR   EMBL; D32137; BAA06863.1; -.
DR   MGD; MGI:106661; Ovgp1.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR001579; Glyco_hydro_18AS.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   ProDom; PD000471; Glyco_hydro_18; 1.
DR   SMART; SM00636; Glyco_18; 1.
DR   PROSITE; PS01095; CHITINASE_18; FALSE_NEG.
KW   Glycoprotein; Fertilization; Repeat; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    721
FT   DOMAIN      486    632       OVIDUCT-SPECIFIC GLYCOPROTEIN.
FT                                21 X 7 AA TANDEM REPEATS OF S-K-T-T-
FT                                [TAP]-G-[IV].
FT   CARBOHYD    402    402       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    442    442       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC...)  (POTENTIAL).
SQ   SEQUENCE   721 AA;  78807 MW;  37246C8F0165652 CRC64;

Query Match             5.1%;  Score 126;  DB 1;  Length 721;
Best Local Similarity  18.9%;  Pred. No. 0.28;
Matches    103;  Conservative   55;  Mismatches   185;  Indels   202;  Gaps   22;

Qy    51 FAGHNSAICA----------SSASYINF-------------------------------      68
         ||||   ||| : :    :  || :|
Db   222 FTGHNSPLFSLPEDSKSSAYAMNYWRKLGTPADKLIMGFPTYGRNFYLLKESKNGLQTAS    281

Qy    69 EGPQNPDNYLVLTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEE    128
         :|::::|::  :  :::  ::       |      :
Db   282 MGPASPGKY-------------TKQAGFLAYYEVCSFVQRAKKHWIDYQ-----------    317

Qy   129 VLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV------    183
         ||  |  :|:   |     ||  :||||| |:  : :|||:
Db   318 ------------YVPYAFKGKEWLG--YDDTISFSYKAMYVKREHFGGAMVWTLDMDDVRGTFC    367

Qy   184 ------------------------------------EIKANGKRAD--FTETFESST-------    202
                                             ::|| :| |: ::||:|:|:
Db   368 GNGPFPLVHILNELLVQTESNSTPLPQFWFTSSVNASGPGSENTALTEVLTTDTIKILPP    427

Qy   203 HGEAPA----------EWTTIDADGDGQWLCLSSGQLGWLTAHGGT---NVVAS-------    244
         || ||           ||| :|:|:|:|:| |||| ||||| ::
Db   428 GGEAMTEVHRRYENMTTVPSDGS-------VTPGGTASPRKHAVTPENNTMAAEAKTMST    481

Qy   245 ---FSWNGMALNPDNYLISKDVTGATKV-------KYYYAVNDGPPGDHYAVM-ISKTGTNA    295
            : || :|  ||||||:::|:  :   ||    :  : ||  |      :::||| ::
Db   482 LDFFSKTTGVSKTTTGVSKTTTGISKTTTGVSKATAGISKTIPEISKATAGVSKTTTTGV    541

Qy   296 GDFTVVFEETPNGINKGGARFGLSTEADGAKPQSVWIER-TVDLPAGTKYVAFRHYNCSD    354
         || | |:|| |: |  :|| || ::|: |:|
Db   542 SKTTTGISKTTTGVSK--TTTGISKTTTGVSKITTGVSKTTTGISKTTTTGIS-------    598

Qy   355 LNVILLDDIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETT----------FEEDGVAT    403
           ||::||||:::|:||:|||:|::| :  ::|:||||:
Db   599 ------QTTTGISKTTTDISKTT-TGISKTTPGISKTTPGMTVIVQTQANEAETTAT    648

Qy   404 GNHEYCVEVK------YTAGVSPKE------------CVNVTDPVQFNPVQNLTGSAVGQK    447
         ||:||| : |      ||||||:|
Db   649 MDHQSVTPTEMDTTLFYLKTMTPSEKETSRKKTMVLEKATVSPREMSATPN-------GQS    702

Qy   448 VTLKW 452
         ||:||
Db   703 KTLKW 707

RESULT 7

SLAP_LACBR
ID   SLAP_LACBR     STANDARD;      PRT;    465 AA.
AC   Q05044;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   S-layer protein precursor (Surface layer protein).
OS   Lactobacillus brevis.
OC   Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=1580;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=ATCC 8287 / DSM 20556;
RX   MEDLINE=93054358; PubMed=1429463;
RA   Vidgren G., Palva I., Pakkanen R., Lounatmaa K., Palva A.;
RT   "S-layer protein gene of Lactobacillus brevis: cloning by polymerase
RT   chain reaction and determination of the nucleotide sequence.";
RL   J. Bacteriol. 174:7419-7427(1992).
CC   -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC       of proteins which coat the surface of bacteria.
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a
CC       S-layer with tetragonal symmetry.
CC   -!- SIMILARITY: SOME, TO THE S-LAYER PROTEIN OF L.ACIDOPHILUS.
CC   -------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------
DR   EMBL; Z14250; CAA78618.1; -.
DR   PIR; A47023; A47023.
KW   Signal; Glycoprotein; Cell wall; S-layer.
FT   SIGNAL        1     30
FT   CHAIN        31    465
SQ   SEQUENCE   465 AA;  48159 MW;  2BE24033392E65A2C CRC64;

Query Match             4.9%;      Score 121.5; DB 1;  Length 465;
Best Local Similarity  19.8%;      Pred. No. 0.32;
Matches     99;  Conservative   48;  Mismatches   201;  Indels  153;  Gaps   20;

Qy    13 TTTLSESFEN-GIPASWKTIDA-----DGDGNNWTTTPPPGGTSFAGHNSAICASSASYIN    67

Qy    26 TTASAKSYATAGAVSTLKTDAATRNVEATGTNALYTKP--GTV-----KGAKVVASKATMAK    80

Db    68 FEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLE    127

Db    81 LASSKKSADYFRAYGVKTNRGSVYYRVVTMDGKYRGVYYG-----GKSDTAFAGGI--    132

Qy   128 EVLTAKTVVTAPEAIRGTRVQGTWYQKTVQ-----LPAGTKYVAFRHFGCTDFFWINLDD   182

Qy   133 ----KSAETTTKADMPARTTGFYLTDTSKNTLWTAPKYTQYXASK                173

Qy   183 VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDG-QGWLCLSSGQLGWLTAHGGTNV   241

Qy   174 VSLYGVAKDTKFTVDQAATKTREGSLYHVTATNGSISGWIYAGK--GFSTTATGTQV    230

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKYKYYYAVNDGPPGDHYAVMISKTGTNAGDFTVV   301

Db   231 LGGLSTDK-------SVTATNDNSVKIVYRTTD-------GTQVGSNTWV            266

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYV-----APRHYNCS    353

Db   267 --------TSTDGTKAGSKVSDKAADQTALEAYINANKPSGYTVTNPNAA          308

Qy   354 DLNY------------ILLDDIQFTMGGSPTPTDYT----------YTVY    381

Db   309 DATYGNTVATVSQAATSKVALKVSGTPVTTALTTADANDKVAANDTTANGSSVAGSTVY    368

Qy   382 RDGTKIKEGLTETTFE---------------EDGVATGNHEYCVEV------KYTAGVSPK 421
```

```
Db    369 AAGTKLAQLTTDLTGEKGQVVTLTAIDTDLEDATFTGTTTYSDLGKAYHYTYTYNKDSA 428
              |||.|  :|.|   |                          ||  .   : ||
Qy    422 BCVNVTVDPVQFNPVQNLTGS 442
              ::|  |:|:|   |:|:|
Db    429 ASSNAS---TQFG--SNVTGT 444
```

RESULT 8
GUXB_CELFI
ID   GUXB_CELFI   STANDARD;   PRT;   1090 AA.
AC   P50899;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Exoglucanase B precursor (EC 3.2.1.91) (Exocellobiohydrolase B)
DE   (1,4-beta-cellobiohydrolase B) (CBP120).
GN   CBHB OR CENE.
OS   Cellulomonas fimi.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Micrococcineae; Cellulomonadaceae; Cellulomonas.
OX   NCBI_TaxID=1708;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
RC   STRAIN=ATCC 484;
RX   MEDLINE=96003898; PubMed=7575482;
RA   Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT   "Cellobiohydrolase B, a second exo-cellobiohydrolase from the
RT   cellulolytic bacterium Cellulomonas fimi.";
RL   Biochem. J. 311:67-74(1995).
RN   [2]
RP   SEQUENCE OF 54-75.
RX   MEDLINE=93209933; PubMed=8458833;
RA   Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT   "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RT   D (CenD), a family A beta-1,4-glucanase.";
RL   J. Bacteriol. 175:1910-1918(1993).
RN   [3]
RP   SEQUENCE OF 54-78.
RX   MEDLINE=94197708; PubMed=8147863;
RA   Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.,
RA   Warren R.A.J., Miller R.C. Jr.;
RT   "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi
RT   CenE, a member of a new family of beta-1,4-glucanases.";
RL   Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
CC   -!- FUNCTION: Hydrolyze cellohexaose to a mixture of cellotetraose,
CC   cellotriose and cellobiose, with only a trace of glucose. It
CC   hydrolyzed cellopentaose to cellotriose and cellobiose, and
CC   cellotetraose to cellobiose, but it did not hydrolyze cellotriose.
CC   Has also weak endoglucanase activity. Hydrolyzes glucosidic bonds
CC   with inversion of anomeric configuration.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC   in cellulose and cellotetraose, releasing cellobiose from the non-
CC   reducing ends of the chains.
CC   -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC   -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC   domain.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL
CC   HYDROLASES).
CC   --
DR   EMBL; L38827; AAB00822.1; -.
DR   PIR; S59077; S59077.
DR   HSSP; P07986; 1BXG.
DR   InterPro; IPR001919; Bac_celose-bind.
DR   InterPro; IPR008965; Cellul_bind.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR000556; Glyco_hydro_48.
DR   InterPro; IPR008928; Glyco_trans_6hp.
DR   Pfam; PF00553; CBM_2; 1.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF02011; Glyco_hydro_48; 1.
DR   PRINTS; PR00844; GLHYDRLASE48.
DR   ProDom; PD011903; Glyco_hydro_48; 1.
DR   SMART; SM06637; CBD_II; 1.
DR   SMART; SM00060; FN3; 3.
DR   PROSITE; PS00561; CBD_BACTERIAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT   SIGNAL   1   33   POTENTIAL.
FT   PROPEP   34   53
FT   CHAIN   54   1090   EXOGLUCANASE B.
FT   DOMAIN   54   699   CATALYTIC (BY SIMILARITY).
FT   DOMAIN   700   785   FIBRONECTIN TYPE-III 1.
FT   DOMAIN   794   884   FIBRONECTIN TYPE-III 2.
FT   DOMAIN   891   978   FIBRONECTIN TYPE-III 3.
FT   DOMAIN   989   1090   CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE   513   513   NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   990   1089   BY SIMILARITY.
SQ   SEQUENCE   1090 AA; 114829 MW; 046BB9D956F2F399 CRC64;

Query Match   4.9%; Score 121; DB 1; Length 1090;
Best Local Similarity   22.3%; Pred. No. 1.1;
Matches 106; Conservative 53; Mismatches 197; Indels 120; Gaps 26;

Qy   26 ASWKT---IDADGDGNNWTTTPPGG-------TSFAGHNSAICASSASYINFBGPCNPD 75
             |||| || ::|:
Db   515 ASWKVPSELKWTGKPDTWNAAAPTGNPGLTVEVTSY-GQDVGVAADTARALLFYAAKSGD 573

Qy   76 NYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTV 135
                                     |:|
Db   574 TASRDKAKALLD---AIWANNQDP-----LGVSAVETRGDYKRFDDTYVAN----GDG 619

Qy   136 VTAPEAIRGTRVQG------------------TWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDD 182
                                              :|:|
Db   620 IYIPSGWTGTMPNGDVIKPGVSFLDIRSFYKKD--PNWSKVQTFLDGGARPQFRYHRFW 676

Qy   183 VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVV 242
           : | || :|
Db   677 AQTAVAGALADYARLFDDGT-------TTPDTTAP-----TVPTG--LQAGVVTSTE 719

Qy   243 ASFSWNGMALNPDNYLISKDV-TGATKVKYYYAVNDGFPGDHYAVMLSKTGTNAGD---F 298
           |:|| :|
Db   720 ATISWT--ASTDDTRVTGYDVRGATKV---------GIATTTSFTDTGLTASTAYAY 766

Qy   299 TVVFEETPNGINKGGARFGLSTEA---DGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDL 355
                                       :|:|:| :|||
Db   767 TVRAFDAAGNVSAPSAALTVTTKATPSDTTAP-------SVPAITS-----SSTA 810

Qy   356 NYILLDDIQFTMGGSPTPTDYT-------YTVYRDGTKIKEGLTETFEEDGV-ATGNHE 407
                :   ::: |::|:::|:
Db   811 NSV------TIGWSAS-TDNAGGSGLAGYDVRGATRVAQ-TTALLFTDTGLTASTAYE 861

Qy   408 YCVEVKYTAG--VSPKECVNVTV-----DPVQFNPVQNLTGSAVGQ-KVTLKWDA 454
           :|:|:|:   :|||:: |:| |:| |
Db   862 YTVRAARDVAGNVSAPSTAVSVTTKSDTTPDTTAPSVPAGLAAMTVTETSVALTWNA 917

RESULT 9
C5AA_BACUD
ID   C5AA_BACUD   STANDARD;   PRT;   1385 AA.
AC   Q45760;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Pesticidial crystal protein crySAa (Insecticidal delta-endotoxin
DE   CryVA(a)) (Crystalline entomocidal protoxin) (152 kDa crystal protein).
GN   CRY5AA OR CRYVA(A) OR CRYVA.
OS   Bacillus thuringiensis (subsp. darmstadiensis).
```

```
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=132264;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL B-18243 / PS17;
RA   Sick A.J., Schwab G.E., Payne J.M.;
RT   "Genes encoding nematode-active toxins cloned from Bacillus
RT   thuringiensis isolate PS17.";
RL   Patent number US5281530, 25-JAN-1994.
CC   -!- FUNCTION: ENDOTOXIN WITH NEMATICIDAL ACTIVITY.
CC   -!- DEVELOPMENTAL STAGE: The crystal protein is produced during
CC       sporulation and is accumulated both as an inclusion and as part of
CC       the spore coat.
CC   -!- MISCELLANEOUS: Toxic segment of the protein is located in the N-
CC       terminus.
CC   -!- SIMILARITY: Belongs to the delta endotoxin family.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L07025; AAA67694.1; -.
DR   PIR; T18213; T18213.
DR   InterPro; IPR005638; endotoxin_C.
DR   InterPro; IPR005639; endotoxin_N.
DR   InterPro; IPR008979; Gal_bind_like.
DR   Pfam; PF03944; endotoxin_C; 1.
DR   Pfam; PF03945; endotoxin_N; 1.
KW   Toxin; Sporulation.
SQ   SEQUENCE 1385 AA; 152439 MW; E29FF11FC799DE95 CRC64;

Query Match    4.8%;   Score 118.5;  DB 1;  Length 1385;
Best Local Similarity   20.1%;   Pred. No. 2.2;
Matches  120;  Conservative  66;  Mismatches  192;  Indels  219;  Gaps  30;

Qy     5 PNPNPNPGTTLSESFENGI---PASWKTIDADGDGNWWTTPPPG---GTSF------ 51
             :  ||      ||    |  ||                :      |||
Db   413 PAVDFQAGPNYVSIDSSNPIIQIMDTWKTPPQGASG--WNTNLMRGSVSGLSFLQRDGT 470

Qy    52 -----AGHNSAICASSA------SYINFEGPQNPDNY----------LVTP-ELSLPN-- 87
                 ||         :|       ||               |  |    |||
Db   471 RLSAGMGGGFADTIYSLPATHYLSYLYGTPYQTSDNYSGHVGALVGVSTPQEATUPNIIG 530

Qy    88 ------------GGTLT-FWVCAQDA----------NYASEHYAVYAS 112
             |                                 |:
Db   531 QPDEQGNVSTMGFPPEKASYGGTVVKEWLNGANAMKLSPCQSIGIPITNVTSGEYQIRCR 590

Qy   113 STGNDASNF------ANALLEEVLTAKTVVTAPEAIRGTRVQG-----------TWYQ 153
              |||                 :: :::  | |||    :  |           ||
Db   591 YASNDNTNVFFNVDTGGANPIFQQINFASTV-----DNNTGVQGANGYVVVKSIATTDN 644

Qy   154 KTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGKRADFTETFESSTHGEAPAEWTTI 213
            |:|| |||   | |||   |  ||    :  |   :  |  ||   : |
Db   645 SFTEIPAKTINVHLTNQGSSDVF---LDRIEFIPFSLPLIYHGSYNTSS-GADDVLW--- 697

Qy   214 DADGDGQGWLCLSSGQLGWLTAHGGTNVVASFSWNGMALNPDNYLISK---------DVTGA 266
            |                     :  ::  ::  :
Db   698 --------------SSSNMNYY-----DIIVNGQANSSSIASSMHILNKGVIKTIDIPGH 739

Qy   267 TKVKY-YYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEADGA 325
              ::  || |                :  |:    |                |
Db   740 SETFFATFPVPEGF---NEVRILAGLPEVSGNITVQSNNPPQPSNNGGGDGGGNGGGDGG 796

Qy   326 KPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPT--------- 374
            |:||                : : :  |:   ||      :| :|
Db   797 Q----------------YN---------FSLSGSDHTTIYHGKLETGIH 820

Qy   375 ---DYTYT---------VYRDGTKIKEGLTETTF------EED---------GVATGNHE 407
            |||                :|      :|             :
```

```
Db   821 VQGNYTYTGTPVLILNAYRNNTVVSSIPVYSPFDITIQEADSLELELQPRYGFATVNGT 880
Qy   408 YCVE---VKYTAGVS-PKECVNVT--VDPVQFNPVQNLTGSAVG----QKVTLKWDA 454
            ::  |: |      ||    : :: | | |||   :  | || ::|| :: ||||
Db   881 ATVKSPNVNYDRSFKLPIDLQNITTQVNALFASGTQMLAHNVSDHDIEEVVLKVDA 937

RESULT 10
GUXA_CELFI
ID   GUXA_CELFI     STANDARD;       PRT;        872 AA.
AC   P50401;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Exoglucanase A precursor (EC 3.2.1.91) (Exocellobiohydrolase A)
DE   (1,4-beta-cellobiohydrolase A) (CBP95).
GN   CBHA.
OS   Cellulomonas fimi.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Micrococcineae; Cellulomonadaceae; Cellulomonas.
OX   NCBI_TaxID=1708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 484;
RX   MEDLINE=94344030; PubMed=8065260;
RA   Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,
RA   Miller R.C. Jr.;
RT   "Cellobiohydrolase A (CbhA) from the cellulolytic bacterium
RT   Cellulomonas fimi is a beta-1,4-exocellobiohydrolase analogous to
RT   Trichoderma reesei CBH II.";
RL   Mol. Microbiol. 12:413-422(1994).
RN   [2]
RP   SEQUENCE OF 41-58.
RX   MEDLINE=93209933; PubMed=8458833;
RA   Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT   "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RT   D (CenD), a family A beta-1,4-glucanase.";
RL   J. Bacteriol. 175:1910-1918(1993).
CC   -!- FUNCTION: This enzyme hydrolyzes 1,4-beta-D-glucosidic linkages of
CC       cellulose. Weak activity against carboxymethylcellulose, bacterial
CC       microcrystalline cellulose and barley beta-glucan. Has also weak
CC       endoglucanase activity. Hydrolyzes glucosidic bonds with inversion
CC       of anomeric configuration.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC   -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC       domain.
CC   -!- SIMILARITY: Belongs to cellulase family B (family 6 of glycosyl
CC       hydrolases).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L25809; AAC36898.1; -.
DR   PIR; S49541; S49541.
DR   HSSP; P07986; 1EXG.
DR   InterPro; IPR001919; Bac_celose-bind.
DR   InterPro; IPR008965; Cellul_bind.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR001524; Glyco_hydro_6.
DR   Pfam; PF00553; fn3; 3.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF01341; Glyco_hydro_6; 1.
DR   PRINTS; PR00014; FNTYPEIII.
```

```
FT   ACT_SITE    481    481       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    547    547       NUCLEOPHILE (BY SIMILARITY).
SQ   SEQUENCE   1034 AA;   118673 MW;   38644C9A649415E9 CRC64;

Query Match             4.5%;    Score 112.5;   DB 1;   Length 1034;
Best Local Similarity  19.9%;    Pred. No. 3.9;
Matches 114;  Conservative 70;  Mismatches 169;  Indels 221;  Gaps 35;

Qy    12 GT-TTLSESFENGIPAS---WK--TIDA-----DGDGNN------WTTPPPGGTSFAGHN 55
             :      |  :                                  :
Db   432 GTWTYLQEGEQKAVPGSKPEWKENVLDRCRSMYERDKNHPSIIWSL---GNESFGGEN 487

Qy    56 SAIC----ASSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAV 109
          :  |
Db   488 FQHMYTFFKEKDSTRLVHYEGIFHHRDYDASDIES------TMYVKPADV---ERYAL 536

Qy   110 ------------YASSTGNDASNFAN---ALLEEVLTAKTVVTAPEAIRGTRVQG---TWY 152
                        :    :   ::        |
Db   537 MNPKKPYILCEYSHAMGNSCGNLYKYWELFDQY----------PILQGGFIWDWK 581

Qy   153 QKTVQLPA--GTKYVAF-RHFGCT----DFFWINLDDVEIKANGKRADFTETFESSTHGE 205
             :  :|   |            :   :       :  :|
Db   582 DQALQATAEDGTSYLAYGGDFGDTPNDGNFCGNGLIFADGTASPKIAEVKKCYQ------ 635

Qy   206 APAEWTTIDADGDGQGWLCLSSGQLWLTAHGGTNVVA-SFSW----NG----------MA 251
Db   636 -PVKWTAVDP------AKGKPAVQNKHLFTNLNAYDFVWTVEKNGELVEKHASLLN 684

Qy   252 LNPD----------------------NYLISKDVTGAT--KVKY--------------- 271
Db   685 VAPDGTDELITLSYPLYEQENETDEFVILTSLRLSKDTAWASAGYEVAYEQFVLPAKAAMP 744

Qy   272 -YYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGLSTE-------ADGA 325
Db   745 SVKAAHPALTVDQNEQTLTVTGTN-----FTAIFDK------RKGQFISYNYERTELLASGF 795

Qy   326 KPQSVWIERTVDLPAGTKYVAFRHYNCS-------------DLNYILLDDIQFTM 367
Db   796 RP-NFW-RAVTDNDLGNKL----HERCQTWRQASLEQHVKKVTVQPQVDFVII-SVELAL 848

Qy   368 GGSPTPTDYTYTVVRDG-TKIKEGL-TETTFE---------------------------- 397
Db   849 DNSLASCVYTYLYNDGEMKIEQSLAPSETMPEIPSIGMLFTMNAAFDSLTWYGRGPHEN 908

Qy   398 ---EDGVATGNHEYCVEVKYTAGVSPKECVNVT 427
Db   909 YWDRKTGAKLALHKGSVKEQVTPYLRPQEGCNKT 942
```

```
Qy   122 ANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
             ||||  ::|  :||| |
Db  1572 ------VITATVDVDNNGFPVKGVTVNFTSNAATAEMTNGGQAVTNEQGKATVTYTNTRS 1624

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQ-GWLCLSSGQLGWLTAHGGTN 240
             ||:  ::: :
Db  1625 SIE---SGARPD--TVEASLENGSSTLSTSINVNADASTAHLTLLQALFDTVSAGDTTN 1678

Qy   241 V---VASFGWNG------MALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISK 290
             ::
Db  1679 LYIEVKDNYGNGVPQQEVTLSVSP-----SEGVTPSNNAIY----TTNHDGNFYA---SF 1726

Qy   291 TGTNAGDFTTV-------FEETPNGI-NKGGARFGLSTEADGAKPQSVWIERTVDLPAG 341
                  ||| |
Db  1727 TATKAGVYQVTATLENGDSMQQTVTYVPNVANARISLAASKDPV------IANNNDLTTL 1780

Qy   342 TKYVAFRHYNC---SDLNYILLDDIQ--FTMGGSPTPTDYTYTVVRDGTKI---KEGLTE 393
             :  |:  :  |   |   :   || |    |   |  :   ||| | :
Db  1781 TATVADTEGNAIANSEVTFTLPEDVRANFTLG-----------DGGKVVTDTEGKAK 1826

Qy   394 TTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLT 440
Db  1827 VTLK--GTKAGAH--TVTASMAGGKSEQLVNFIADTLTAQVNLNVT 1869
```

```
RESULT 12
BGAL_BACME
ID   BGAL_BACME    STANDARD;      PRT;   1034 AA.
AC   O52847;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Beta-galactosidase (EC 3.2.1.23) (Lactase).
GN   BGAM.
OS   Bacillus megaterium.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1404;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 319;
RA   Strey J.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC       galactose residues in beta-D-galactosides.
CC   -!- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; AJ000733; CAA04267.1; -.
DR   PIR; T30574; T30574.
DR   HSSP; P00722; 1BGL.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR006101; Glyco_hydro_2.
DR   InterPro; IPR006102; Glyco_hydro_2Ig.
DR   InterPro; IPR006104; Glyco_hydro_2SB.
DR   InterPro; IPR006103; Glyco_hydro_2TIM.
DR   InterPro; IPR004200; Glyco_hydro_42C.
DR   InterPro; IPR004199; Glyco_hydro_42N.
DR   Pfam; PF02930; Bgal_small_C; 1.
DR   Pfam; PF02929; Bgal_small_N; 1.
DR   Pfam; PF00703; Glyco_hydro_2; 1.
DR   Pfam; PF02836; Glyco_hydro_2_C; 1.
DR   Pfam; PF02837; Glyco_hydro_2_N; 1.
DR   PRINTS; PR00132; GLHYDRLASE2.
DR   PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR   PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KW   Hydrolase; Glycosidase.
```

```
RESULT 13
ALS1_CANAL
ID   ALS1_CANAL    STANDARD;      PRT;   1260 AA.
AC   P46590;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Agglutinin-like protein 1 precursor.
GN   ALS1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11651 / B792;
RX   MEDLINE=95272392; PubMed=7752895;
RA   Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RT   "Candida albicans ALS1: domains related to a Saccharomyces cerevisiae
RT   sexual agglutinin separated by a repeating motif.";
RL   Mol. Microbiol. 15:39-54 (1995).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-glycosylated and O-glycosylated (Potential).
CC   -!- SIMILARITY: TO YEAST SAG1.
```

```
Db   566  --------------PGG?DSVIIREPPNHTVTTTEYW------------------SQSYA 593

Qy   200  SSTHGEAPAEWTTIDADGDGQGWICLSSGQLGWLTAHGGTNVVASFSWNGMALNPDNYLI 259
                                                            ::|:  :::
Db   594  TTTTVTAPP-----------------------------------------GTDTV---- 618
                                                              |||
Qy   260  SKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPN 307
                                                    |||||
Db   619  --------TTTEYW-----------SQSFATTTTVTGPPSGTDTVIIREPPN 651


RESULT 14
ADP1_MYCGA
ID   ADP1_MYCGA    STANDARD;      PRT;   1122 AA.
AC   Q49379; Q49437; Q53351;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Adhesin P1 precursor (Cytadhesin P1) (Attachment protein).
GN   GAPA OR MGC1.
OS   Mycoplasma gallisepticum.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2096;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S6;
RX   MEDLINE=96201559; PubMed=8613358;
RA   Keeler C.L. Jr., Hnatow L.L., Whetzel P.L., Dohms J.E.;
RT   "Cloning and characterization of a putative cytadhesin gene (mgc1)
RT   from Mycoplasma gallisepticum.";
RL   Infect. Immun. 64:1541-1547(1996).
RN   [2]
RP   SEQUENCE OF 1-12 FROM N.A.
RC   STRAIN=S6;
RA   Hnatow L.L., Keeler C.L. Jr., Tessmer L., Dohms J.E.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 378-570 FROM N.A.
RC   STRAIN=S6;
RX   MEDLINE=93371270; PubMed=8363503;
RA   Dohms J.E., Hnatow L.L., Whetzel P., Morgan R., Keeler C.L. Jr.;
RT   "Identification of the putative cytadhesin gene of Mycoplasma
RT   gallisepticum and its use as a DNA probe.";
RL   Avian Dis. 37:380-388(1993).
RN   [4]
RP   SEQUENCE OF 159-1122 FROM N.A.
RC   STRAIN=S6;
RA   Goh M.S., Geary S.J.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Could be involved in cytadherence.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the adhesin P1 family.
CC   -!- CAUTION: Ref.4 sequence differs from that shown due to a
CC       frameshift in position 159 to 213.
CC   ----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------
DR   EMBL; U34842; AAB02987.1; -.
DR   EMBL; U44804; AAC83385.1; ALT_FRAME.
DR   PIR; T18346; T18346.
KW   Cytadherence; Signal; Transmembrane.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1122       ADHESIN P1.
FT   TRANSMEM    997   1021       POTENTIAL.
FT   DOMAIN     1002   1009       POLY-ILE.
FT   CONFLICT    313    314       DM -> IW (IN REF. 4).
FT   CONFLICT    382    382       Y -> F (IN REF. 3).
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------
DR   EMBL; L25902; AAC41649.2; -.
DR   InterPro; IPR008840; Candida_ALS.
DR   Pfam; PF05792; Candida_ALS; 1.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1260       AGGLUTININ-LIKE PROTEIN 1.
FT   DOMAIN      433    792       10 X 36 AA TANDEM REPEATS.
FT   REPEAT      433    468       1-1.
FT   REPEAT      469    504       1-2.
FT   REPEAT      505    540       1-3.
FT   REPEAT      541    576       1-4.
FT   REPEAT      577    612       1-5.
FT   REPEAT      613    648       1-6.
FT   REPEAT      649    684       1-7.
FT   REPEAT      685    720       1-8.
FT   REPEAT      721    756       1-9.
FT   REPEAT      757    792       1-10.
FT   DOMAIN      983   1152       2 X 26 AA APPROXIMATE REPEATS.
FT   REPEAT      983   1043       2-1.
FT   REPEAT     1092   1152       2-2.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      450    455       POLY-THR.
FT   DOMAIN      486    491       POLY-THR.
FT   DOMAIN      522    527       POLY-THR.
FT   DOMAIN      558    563       POLY-THR.
FT   DOMAIN      594    599       POLY-THR.
FT   DOMAIN      630    635       POLY-THR.
FT   DOMAIN      666    671       POLY-THR.
FT   DOMAIN      702    707       POLY-THR.
FT   DOMAIN      738    743       POLY-THR.
FT   DOMAIN      774    779       POLY-SER.
FT   DOMAIN      874    877       POLY-THR.
FT   CARBOHYD    471    471       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    723    723       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    820    820       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    886    886       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    918    918       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    973    973       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD   1045   1045       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD   1068   1068       N-LINKED (GLCNAC...)    (POTENTIAL).
SQ   SEQUENCE   1260 AA;  132641 MW;  763D1063A2354C24 CRC64;

Query Match             4.5%;  Score 112.5;  DB 1;  Length 1260;
Best Local Similarity  19.5%;  Pred. No. 5;
Matches    68;  Conservative   30;  Mismatches   77;  Indels  173;  Gaps   16;

Qy     5  PNPNPNPGTTTL-SESPENGIPASWKTIDADGDGNN---------------------W---- 40
             ||||:|:|:||
Db   432  PSPNPTVSTTEYWSQSF----ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFAT 486

Qy    41  --TTTPPGGTSFAGHNSAICASSASYINFEGQNP--------DNYLVTPELSLPNGG 89
            ||||||          ||        ::  |||
Db   487  TTTTVTAPPGGT-----DSVI-----IREPPNPTVTTTEYWSQSFATTTTVTAPPGG 532

Qy    90  TLTFWV------CAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIR 143
             :|:|                 ::|||
Db   533  TDSVIIREPPNPTVTTTEYWSQSYA------------------TTTTVTAP--- 565

Qy   144  GTRVQGTWYQKTVQLPAGTKVVAFR----HPGCTDFFWINLDDVEIKANGKRADFTETFE 199
                |||                          ::|        ::|:
```

```
FT   CONFLICT     501    501       I -> T (IN REF. 3).
FT   CONFLICT     568    568       R -> G (IN REF. 3).
FT   CONFLICT     570    570       T -> A (IN REF. 3).
FT   CONFLICT     693    695       DIL -> VVT (IN REF. 4).
SQ   SEQUENCE    1122 AA;  121305 MW;  155C34DA2D6C3C65 CRC64;

Query Match       4.5%;  Score 110.5;  DB 1;  Length 1122;
Best Local Similarity  19.0%;  Pred. No. 5.9;
Matches    94;  Conservative  77;  Mismatches  193;  Indels  131;  Gaps  26;

Qy     42 TTPPGGTSPAGHNSAICASSASYINFBEGQNPDNYLVTPELSLPNGGTLTFWVCAQDAN  101
                 :  : ||                   |:|: :    :
Db    437 TAEAPGNTKTVGYPYGILLSA---ISFDATR-----------NGLALAAPALGQDVG  479

Qy    102 YASBHYA---------------VYASST---GNDASNFAN------ALLEEVLTAKTV  135
              :    :||:               |:|: :                ||
Db    480 Y---HFVPRLAVGGVSSPRGANGNIFLGSAITWGTNGGNFLDTKWHSPAVIEDAPTTFIT  536

Qy    136 VTAPEAIRGTRVQGTWYQKTVQLP-AGTKYVAFRHFGCTDFFWINLDDVEIK--ANGK  190
                 ||| |  ||  |  :                 |:|  :        :
Db    537 VNSSGVLQNS--GSQQSTSTPMPNSNGNESIPYRWTNSYDYNSVRFAALISKPAGGNTK  593

Qy    191 RAD--FTETFESSTHGEAPAEWTTID------ADGDGQGWLCLSSGQLGWLTAHGGTNVV  242
              :|:        |||                :||||  |||  |:||:  ||
Db    594 QVESLFTTALKLDTLNSLPNKFTQENNIFFSYAMLDGRQWSLGTRKDSTWLT---TNTI  649

Qy    243 ASFSWN------------GMALNPDNYLIS----------KDV----TGATKVKYYAV  275
                 |:           :| : : :|          ||:        ||:|:|:
Db    650 NNFYNTQQQLASTAAGENANPRNILNALTTAKGFDRRDIGNVDILYSNNTNKFTYYQV  709

Qy    276 ND------------GFPGDHYAVMISKTGTNAGDFTVVPEETPNGINK-GGARFGLSTEADG  324
                         :|: :     ||:| ||||      ||:  |:|| ||||||||||:
Db    710 GGAITTWPEVQVNYKTSANITYYNLTRTDFGSTTPATQDANTVSSKLNGAYLSSTGDQQG  769

Qy    325 AKPQSVWIERTVDLPAGTKYV--AFRHYNCSDLNVILLDDIQFTMGGSPTPTDYTYTVR  382
               |:  :||||: : |||:|:    :|:  : :   :| :|  |  |  || ||::|||
Db    770 WYNGSIYYKKASFTPSSQGYTWQDFKGLITTASNAVISN---WTKAG------YSIRP  818

Qy    383 DGTKIKEGLTETTFEEDGVATGN---HEYCVE---VKYTAGVSPKECVNVTVDPVQ  432
               ||::|  |  |::|||  | |     |:||    ||:|||||||||:| |||
Db    819 DDDTV-FSVSKIPFEKEITAAVNVRSLDSYYVQLNGETSVNTVARVSPDSSA-ITLNPKR  876

Qy    433 F-NPVQNLTGSAVGQ  446
           | ||  :|:||||||
Db    877 ITNPLMN-RDNVIGQ  890
```

```
RESULT 15
PME_ASPAC
ID   PME_ASPAC       STANDARD;    PRT;   331 AA.
AC   Q12535;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Pectinesterase precursor (EC 3.1.1.11) (Pectin methylesterase) (PE).
GN   PME1.
OS   Aspergillus aculeatus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5053;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KSM 510;
RX   MEDLINE=97079238; PubMed=8920970;
RA   Christgau S., Kofod L.V., Halkier T., Andersen L.N., Hockauf M.,
RA   Dorreich K., Dalboege H., Kauppinen S.;
RT   "Pectin methyl esterase from Aspergillus aculeatus: expression
RT   cloning in yeast and characterization of the recombinant enzyme.";
RL   Biochem. J. 319:705-712(1996).
CC   -!- FUNCTION: Involved in maceration and soft-rotting of plant tissue.
CC   -!- CATALYTIC ACTIVITY: Pectin + N H(2)O = N methanol + pectate.
CC   -!- SIMILARITY: Belongs to the pectinesterase family.
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U49378; AAB42153.1; -.
DR   InterPro; IPR000070; Pectinesterase.
DR   Pfam; PF01095; Pectinesterase; 1.
DR   PROSITE; PS00800; PECTINESTERASE_1; 1.
DR   PROSITE; PS00503; PECTINESTERASE_2; 1.
KW   Hydrolase; Aspartyl esterase; Cell wall; Signal.
FT   SIGNAL        1     17               BY SIMILARITY.
FT   CHAIN        18    331               PECTINESTERASE.
FT   ACT_SITE    162    162               BY SIMILARITY.
FT   ACT_SITE    183    183               BY SIMILARITY.
SQ   SEQUENCE   331 AA;  35681 MW;  1F1C81BF1E32174F CRC64;

Query Match       4.4%;  Score 109.5;  DB 1;  Length 331;
Best Local Similarity  22.5%;  Pred. No. 1.4;
Matches    69;  Conservative  33;  Mismatches  99;  Indels  105;  Gaps  15;

Qy    108 AVYASSTGNDASNFANALLEVLIAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGT-KYVA  166
              |  :|  |: ||  |  ::|::  |:    |  || :||||    |  |  ||:
Db     27 AIVVAKSGGDYTTIGDAI--DALSTSTTDQTIFIE---EGT-YDEQVLPAMTGKVII   79

Qy    167 FRHFGCTDFFWINL---DDVEIKANGKRADFTETFE-----------SSTHGEAPAE  209
            :|:|  ||| |:         :|  :|   ||| |             :|  |||
Db     80 YGQTENTDSYADNLVTITHAISYEDAGESDLTATFRNKAVGSQVYNLNIANTCGQACHQ  139

Qy    210 WTTIDADGDGQGWL-CLSSGQLGWLTAHGGTNVV-----------------------  242
            |||| | ||||||  |:||||||||||||:||
Db    140 ALALSAWADQQGYYGCNFTGYQDTLLAQTGNQLYINSYIEGAVDFIFGQHARAWFQNVDI  199

Qy    243 ------ASFSWNGMALNPDN--YLISKDVTGA-------TKVKYYYAVNDGFPGDHYAV  286
                  :|||||  |  |:|    ||:|: :||         | |:|:|:|||  |:|:|
Db    200 RVVEGPTSASITANGRSSETDTSYYVINKSTVAAKEGDVAEGTYYL---GRPWSEYA-  254

Qy    287 MISKTGTNAGDFTVVPEET--PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKY  344
            ||:|||||||||||:|||   |:| :|||||||||||||| ||:|||:||  |: |||::
Db    255 --------------------RVVFQQTSMTNVINSLG----------WTEWSTSTP-NTEY  284

Qy    345 VAFRHY  350
            |:|:|:
Db    285 VTFGEY  290
```

```
Search completed: May 18, 2004, 11:43:35
Job time : 9.5342 secs
```

```
                GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May 18, 2004, 11:36:20 ; Search time 33.3368 Seconds
                                       (without alignments)
                                       4315.838 Million cell updates/sec

Title:            US-08-570-311-20
Perfect score:    2480
Sequence:         1 GTPNPNPNPNPGTTTLSESF........QNLTGSAVGQKVTLKWDAPN 456

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:         1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:   1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :  SPTREMBL.25:*
        1:  sp_archea:*
        2:  sp_bacteria:*
        3:  sp_fungi:*
        4:  sp_human:*
        5:  sp_invertebrate:*
        6:  sp_mammal:*
        7:  sp_mhc:*
        8:  sp_organelle:*
        9:  sp_phage:*
       10:  sp_plant:*
       11:  sp_rodent:*
       12:  sp_virus:*
       13:  sp_vertebrate:*
       14:  sp_unclassified:*
       15:  sp_rvirus:*
       16:  sp_bacteriap:*
       17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2094	84.4	1723	2	P72197	P72197 porphyromon
2	2082	84.0	1687	2	Q9R9B7	Q9r9b7 porphyromon
3	2082	84.0	1704	2	Q51816	Q51816 porphyromon
4	2079	83.8	1723	2	P72194	P72194 porphyromon
5	2071	83.5	1358	2	P96967	P96967 porphyromon
6	2070	83.5	1706	2	Q51839	Q51839 porphyromon
7	2059	83.0	1706	2	Q51838	Q51838 porphyromon
8	2054.5	82.8	1223	2	Q92NB5	Q92NB5 porphyromon
9	2039.5	82.2	1097	2	P72196	P72196 porphyromon
10	2039.5	82.2	1732	2	O07442	O07442 porphyromon
11	2034.5	82.0	1732	2	Q51817	Q51817 porphyromon
12	2020.5	81.5	1732	2	O52050	O52050 porphyromon
13	682	27.5	925	2	Q9F4J0	Q9F4J0 porphyromon
14	561.5	22.6	312	2	Q9KIB3	Q9KIB3 porphyromon
15	233	9.4	293	2	Q9XBU9	Q9XBU9 porphyromon
16	152.5	6.1	1742	16	Q8P377	Q8P377 xanthomonas
17	144	5.8	3346	16	Q7WN54	Q7WN54 bordetella
18	139.5	5.6	1341	16	Q8UAU1	Q8UAU1 agrobacteri
19	139	5.6	2215	16	Q7WBN0	Q7WBN0 bordetella
20	135	5.4	3988	17	Q8TPZ1	Q8TPZ1 methanosarc
21	134.5	5.4	2768	16	Q8E9G6	Q8E9G6 shewanella
22	134	5.4	1541	16	Q7UQJ9	Q7UQJ9 rhodopirell
23	134	5.4	1800	2	Q9L948	Q9L948 pseudomonas
24	134	5.4	8682	16	Q88RG2	Q88RG2 pseudomonas
25	133.5	5.4	680	2	Q52644	Q52644 ruminococcu
26	133	5.4	2522	16	Q8EKA6	Q8EKA6 shewanella
27	132.5	5.3	555	16	Q826W1	Q826W1 streptomyce
28	132.5	5.3	691	16	Q9RZS7	Q9RZS7 deinococcus
29	132.5	5.3	1115	4	Q9HD43	Q9HD43 homo sapien
30	131.5	5.3	1256	16	Q8NX96	Q8NX96 staphylococ
31	131.5	5.3	1357	17	Q8THC8	Q8THC8 methanosarc
32	131	5.3	2468	16	Q9I2M3	Q9I2M3 pseudomonas
33	131	5.3	2219	16	Q88WI9	Q88WI9 lactobacill
34	130.5	5.3	880	16	Q9RMB8	Q9RMB8 arthrobacte
35	130	5.3	955	17	Q8TQ91	Q8TQ91 methanosarc
36	130	5.3	1386	17	Q8TI72	Q8TI72 methanosarc
37	129.5	5.2	756	9	Q858B6	Q858B6 enterobacte
38	129.5	5.2	1744	16	Q82YW8	Q82YW8 enterococcu
39	129	5.2	7716	16	Q7UWZ8	Q7UWZ8 rhodopirell
40	128.5	5.2	1248	16	Q99V41	Q99V41 staphylococ
41	128.5	5.2	1255	2	Q7WTC6	Q7WTC6 staphylococ
42	128.5	5.2	2656	5	Q9GNU3	Q9GNU3 paracentrot
43	128	5.2	1672	16	Q8Y366	Q8Y366 ralstonia s
44	128	5.2	4936	16	Q8YKJ3	Q8YKJ3 anabaena sp
45	127.5	5.1	1684	2	Q03658	Q03658 unidentifie

ALIGNMENTS

```
RESULT 1
P72197
ID   P72197      PRELIMINARY;      PRT;   1723 AA.
P72197
AC   P72197;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA   Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT   "Molecular cloning and characterization of Porphyromonas gingivalis
RT   Lys-gingipain.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U54691; AAA99810.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25_I.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE 1723 AA;  186831 MW;  4508A7E50197CEBD CRC64;

Query Match             84.4%;    Score 2094;  DB 2; Length 1723;
Best Local Similarity   86.2%;    Pred. No. 6.8e-128;
```

```
Matches 394;  Conservative  17;  Mismatches  40;  Indels   6;  Gaps   4;

Qy      1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA   60
          ||||||||||||||||||||||||||||||||| |||: :  :|:||
Db    969 GTPNPNPNPNPG-TTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 1025

Qy     61 SSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVIASSTGNDAS  119
          :: : ||||:|  |||:| |: :|| |||||||||||||||||||||||||||||||||
Db   1026 YSESFGLGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVIASSTGNDAS  1085

Qy    120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN  179
          ||  :|||: |:||:|:| ::|||||||||||||||||||||||||||||||||
Db   1086 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 1143

Qy    180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGT  239
          |:|||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db   1144 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGT 1203

Qy    240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 1263

Qy    300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL  359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1264 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 1323

Qy    360 LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  419
          |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 1383

Qy    420 PKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN                         456
          |||:|||||||:|| |||| ||:||:
Db   1384 PKKCVNVTINPTQFNPVKNLKAQPDGDGDVVLKWEAPS                        1420
```

RESULT 2
Q9R9B7
```
ID   Q9R9B7;              PRELIMINARY;      PRT;  1687 AA.
AC   Q9R9B7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin/protease.
GN   HAGE.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Dong H., Progulske-Fox A.;
RT   "Cloning and characterization of hagE from P. gingivalis 381.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF026946; AAD01810.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE  1687 AA;  183702 MW;  D085B516A399FE70 CRC64;
```

```
Query Match             84.0%;        Score 2082;  DB 2;  Length 1687;
Best Local Similarity   86.2%;        Pred. No. 4e-127;
Matches 392;  Conservative  17;  Mismatches  40;  Indels   6;  Gaps   4;

Qy      3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS   62
          ||||||||||||||||||||||||||||||||| |||: :  :|:|
Db    935 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS  991

Qy     63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF  121
          :: ||||:|  |||:| |: :|| ||||: ||||||||||||||||||||||||||||||
Db    992 ESFGLGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNF  1051

Qy    122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD  181
          |:|||: |:||:|:| ::|||||||:|||||||||||||||||||||||||:::|:|||
Db   1052 TNALLEETITAKG--VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy    182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGTNV  241
          :|||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db   1110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1169

Qy    242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1229

Qy    302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1289

Qy    362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1349

Qy    422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN                           456
          |||||||:|| |||| ||: |:|
Db   1350 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS                          1384
```

RESULT 3
Q51816
```
ID   Q51816;              PRELIMINARY;      PRT;  1704 AA.
AC   Q51816;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arg-gingipain-1 proteinase.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95138080; PubMed=7836351;
RA   Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
RA   Travis J., Barr P.J.;
RT   "Molecular cloning and structural characterization of the Arg-
RT   gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a
RT   proteinase-adhesin polyprotein.";
RL   J. Biol. Chem. 270:1007-1010(1995).
DR   EMBL; U15282; AAA69539.1; -.
DR   PIR; A55426; A55426.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
```

```
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25;_1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   CHAIN         228    719       MATURE 50-KDA CYSTEINE PROTEINASE
FT                                  GINGIPAIN.
SQ   SEQUENCE   1704 AA;  185436 MW;  6A34B40131C2A676 CRC64;

Query Match             84.0%;  Score 2082;  DB 2;  Length 1704;
Best Local Similarity   86.2%;  Pred. No. 4e-127;
Matches  392;  Conservative  17;  Mismatches  40;  Indels  6;  Gaps  4;

Qy     3 PNPNPNPNPGTTLSESEPENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASS   62
         ||||||||||||||||||||||||||||||||||||      :||||::||||:
Db   952 PNPNPNPNPGTTLSESEPENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy    63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF  121
         :|:  ||| :|||:||||||:||||ː||||||||||||||||||||||||||||||
Db  1009 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD  181
         ||||||ː|| | |:||||| :|||:||||||||||||||||||||||::||:||ː:||
Db  1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLTAHGGTNV     241
         :||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db  1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  301
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
         |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db  1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPFDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK    421
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1307 DIQFTMGGSPTPFDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  1366

Qy   422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN                          456
         |||||:|ː| |||||:||: ||ː:|:|:||||ː|:
Db  1367 ECVNVTINPTQFNPVKNLKAQPDGGGDVVLKWEAPS                        1401
```

RESULT 4

```
P72194
ID   P72194;       PRELIMINARY;       PRT;   1723 AA.
AC   P72194;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97044756; PubMed=8889827;
RA   Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
RT   "Cloning and sequencing of the gene encoding a novel lysine-specific
RT   cysteine proteinase (Lys-gingipain) in Porphyromonas gingivalis:
RT   structural relationship with the arginine-specific cysteine proteinase
RT   (Arg-gingipain).";
RL   J. Biochem. 120:398-406(1996).
DR   EMBL; D83258; BAA11870.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
```

```
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1723 AA;  187261 MW;  5628963D251493EB CRC64;

Query Match             83.8%;  Score 2079;  DB 2;  Length 1723;
Best Local Similarity   85.8%;  Pred. No. 6.4e-127;
Matches  392;  Conservative  17;  Mismatches  42;  Indels  6;  Gaps  4;

Qy     1 GTPNPNPNPNPGTTLSESEPENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA    60
         |||||||||||||||||||||||||||||||||||       |:||||::||||
Db   969 GTPNPNPNPNPGTTLSESEPENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 1025

Qy    61 SSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS  119
          :|:  ||| :|||:||||||:||||ː||||||||||||||||||||||||||||||
Db  1026 YSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 1085

Qy   120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN  179
         ||:||||||ː||| | |:||||| :|||:|||||||||||||||||||||::||:||ː:
Db  1086 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 1143

Qy   180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLTAHGGT     239
         || :||||||||||||||||||||||||||||||||||||||||||||||||  |||
Db  1144 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGT 1203

Qy   240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  299
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1204 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 1263

Qy   300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL  359
         |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db  1264 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 1323

Qy   360 LDDIQFTMGGSPTPFDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS    419
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1324 LDDIQFTMGGSPTPFDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  1383

Qy   420 PKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN                        456
         ||:|||||:|ː| |||||:||: ||ː:|:|:||||ː|:
Db  1384 PKVCVNVTINPTQFNPVKNLKAQPDGGGDVVLKWEAPS                      1420
```

RESULT 5

```
P96967
ID   P96967;       PRELIMINARY;       PRT;   1358 AA.
AC   P96967;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin.
GN   HAGD.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Lepine G., Whitlock J., Wojciechowski L., Progulske-Fox A.;
RT   "Cloning, sequencing and characterization of hagD, a member of the
RT   HArep multigene family in Porphyromonas gingivalis.";
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U68468; AAB49691.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
```

```
RL   Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96311339; PubMed=8711096;
RA   Slakeski N., Cleal S.M., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtR that encodes
RT   an arginine-specific thiol proteinase and multiple adhesins.";
RL   Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Reynolds E.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Slakeski N.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L26341; AAC18876.1; -.
DR   HSSP; P95493; 1CVR.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase C25.
DR   InterPro; IPR005536; Peptidase C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN       228   1706       ARGININE-SPECIFIC THIOL PROTEASE.
SQ   SEQUENCE   1706 AA;  185626 MW;  E8BDF07C9813B844 CRC64;

Query Match             83.5%;  Score 2070;  DB 2;  Length 1706;
Best Local Similarity   85.7%;  Pred. No. 2.5e-126;
Matches  390;  Conservative  18;  Mismatches  41;  Indels  6;  Gaps  4;

Qy      3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASS   62
            |||||||||||||||||||||||||||||||||       |||||||||  | |||
Db    954 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWMKPGNAPG---IAGYNSNGCVYS 1010

Qy     63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
            | :  |||||||||||| ||||||| || |  |||||||||||||||||||||||||||
Db   1011 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

Qy    122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
            ||||||||||||||||| ::|||:  |||||||||||||||||||||||||::|||||||
Db   1071 TNALLEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy    182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1129 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1188

Qy    242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db   1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy    302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1249 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy    362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1309 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy    422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

```
Db    1369 KCVNVTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403
             :||||||: ||||||: ||||||  | |||:||

RESULT 7
Q51838
ID   Q51838      PRELIMINARY;    PRT;   1706 AA.
AC   Q51838;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Protease precursor.
GN   PRPR1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96071894; PubMed=7591131;
RA   Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.;
RT   "Characterization, genetic analysis, and expression of a protease
RT   antigen (PrpRI) of Porphyromonas gingivalis W50.";
RL   Infect. Immun. 63:4744-4754(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.;
RT   "The prpR1 and the prR2 arginine-specific protease genes of
RT   Porphyromonas gingivalis W50 produce five biochemically distinct
RT   enzymes.";
RL   Mol. Microbiol. 23:0-0(1997).
DR   EMBL; X82680; CAA57997.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; I.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   SIGNAL.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        228    719       ALPHA-PROTEASE.
FT   CHAIN        720   1262       BETA-ADHESIN.
SQ   SEQUENCE   1706 AA;  185705 MW;  0E56DCD87FDA8CDD CRC64;

Query Match     83.0%;  Score 2059;  DB 2;  Length 1706;
Best Local Similarity   85.5%;  Pred. No. 1.3e-125;
Matches 389;  Conservative 17;  Mismatches 43;  Indels 6;  Gaps 4;

Qy      3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
            |||||||||||||||||||||||||||||||||||||||| ||:  ||
Db    954 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy     63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
            ||| |:|||||||||||||||||||:|||:|||||||||||||||||||||||||||||
Db   1011 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1070

Qy    122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
            ::|||:||||:||||:|||:||||||:||||::||||:|||||||||||||:|||:||:|
Db   1071 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy    182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLWLTAHGGTNV 241
            |||||||||||||||||||||| |||||||||||||||||||||||||  ||||||||||
Db    1129 EVEIKANGKRADFTETFESSTHGEATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1188

Qy    242 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
            ::||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy    302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1249 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy    362 DIQFTMGGSPTPDTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 421
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1309 DIQFTMGGSPTPDTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy    422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
            :|||||  ||:||||:|  |:|: ||:|:||:
Db    1369 KCVNVTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 8
Q9ZNB5
ID   Q9ZNB5      PRELIMINARY;    PRT;   1223 AA.
AC   Q9ZNB5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   130k-HMGD (fragment).
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=99143166; PubMed=9988746;
RA   Shibata Y., Hayakawa M., Takiguchi H., Shiroza T., Abiko Y.;
RT   "Determination and characterization of the hemagglutinin-associated
RT   short motifs found in Porphyromonas gingivalis multiple gene
RT   products.";
RL   J. Biol. Chem. 274:5012-5020(1999).
DR   EMBL; AB019363; BAA34341.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; I.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   1223 AA;  131542 MW;  00225CD2BA9F91B3 CRC64;

Query Match     82.8%;  Score 2054.5;  DB 2;  Length 1223;
Best Local Similarity   85.2%;  Pred. No. 1.6e-125;
Matches 391;  Conservative 15;  Mismatches 44;  Indels 9;  Gaps 6;

Qy      1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
            |||||||||||| :||||||||||||||||||||||||||||||||||||||||||
Db    460 GTPNPNPNPNPCTTLSESFENGIPASWKTIDADGDGHGHGWKPGNAPG---IAGYNSNGC 516

Qy     60 ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
            || :||| |:|||||||||||||||||||||:||||||||||||||||||||||||||
Db    517 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 576

Qy    119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
            :|| ::|||:||||:||||:|||:||||||:||||::||||:|||||||||||||:|||
Db    577 SNFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 634
```

```
Db  391  VYSEFGLGGIVGVLTPDNYLIITPALDLLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA  450
         ||| :|.:||||| ::|:|:|  :|||:|||||||||||||||||  ::|| |||||||||||
Qy  119  SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI    178
         ||| : |||||  : | : |  ::| :  .|| | | |:| |||||||||||:|:| :|:|
Db  451  SNFTNALLEETTIAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI    508

Qy  179  NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGG    238
         :|.:|||||||||||||||||||||||||||||||||||||||||||||||.|||||||
Db  509  DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGG    568

Qy  239  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF  298
         ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  569  SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF  628

Qy  299  TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI  358
         |||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db  629  TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI  688

Qy  359  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV  418
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  689  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV  748

Qy  419  SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP  455
         ||:|||||| :|:||||||||||    |||||||||||
Db  749  SPKKCVNVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP  787
```

```
RESULT 10
ID   O07442      PRELIMINARY;      PRT;     1732 AA.
AC   O07442;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine-specific cysteine proteinase.
GN   PRTK.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=99235907; PubMed=10219167;
RA   Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtK that encodes
RT   a lysine-specific cysteine proteinase and three sequence-related
RT   adhesins.";
RL   Oral Microbiol. Immunol. 14:92-97(1999).
DR   EMBL; U75366; AAB60809.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01164; Peptidase_C25_C; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE  1732 AA;  187914 MW;  45D5B9137739170 CRC64;

Query Match             82.2%;  Score 2039.5;  DB 2;  Length 1732;
Best Local Similarity   84.5%;  Pred. No. 2.4e-124;
Matches   388;  Conservative  18;  Mismatches   44;  Indels   9;  Gaps   6;

Qy    1  GTPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTPPGGTSFAGHNSAIC   59
         |||||||||||  ||||||||||||||||||||||||| ::|| |||||||||||
Db  969  GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGHGWKPGNAPG---IAGYNSNGC 1025
```

```
Db  179  NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGG  238
         :|.:|||||||||||||||||||||||||||||||||||||||||||||||.|||||||
Qy  635  DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGG  694

Db  239  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
         ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  695  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 754

Db  299  TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
         |||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Qy  755  TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 814

Db  359  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  815  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 874

Db  419  SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP  455
         ||:|||||| :|:||||||||||    |||||||||||
Qy  875  SPKVCVNVTINPTQFNPVQNLTAEQAPNSMDAILKWNAP  913
```

```
RESULT 9
ID   P72196      PRELIMINARY;      PRT;     1097 AA.
AC   P72196;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   TonB-linked adhesin precursor.
GN   TLA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97386416; PubMed=9244265;
RA   Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA   Curtis M.A.;
RT   "The tla gene of Porphyromonas gingivalis W50: a homologue of the
RT   arginine-specific protease precursor (PrpRI) which shares sequence
RT   similarity to TonB-linked receptors.";
RL   J. Bacteriol. 179:4778-4788(1997).
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR   EMBL; Y07618; CAA68897.1; -.
DR   GO; GO:0019867; C:outer membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB dep.Rec; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Membrane; Outer membrane; Receptor; Signal; TonB box.
FT   SIGNAL         1     53       POTENTIAL.
SQ   SEQUENCE  1097 AA;  118731 MW;  73BBA337B421F8B9 CRC64;

Query Match             82.2%;  Score 2039.5;  DB 2;  Length 1097;
Best Local Similarity   84.5%;  Pred. No. 1.3e-124;
Matches   388;  Conservative  18;  Mismatches   44;  Indels   9;  Gaps   6;

Qy    1  GTPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTPPGGTSFAGHNSAIC   59
         |||||||||||  ||||||||||||||||||||||||| ::|| |||||||||||
Db  334  GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGHGWKPGHGNAPG---IAGYNSNGC 390

Qy   60  ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
         |:  :  |  |  :|  |||||| | ||||  |||||||||||||||||||||||||||||
Db  ...
```

```
Qy      1  GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDDADGDGNWTTPPPGGTSFAGHNSAIC    59
           |||||||||||||  :|::|||||||||||||||||||||||||||  :||:|||||
Db    969  GTPNPNPNPNPGTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC   1025

Qy     60  ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA   118
           ||:::.  |||::||||||||||||||||||||||||||||||||||||||||||||||
Db   1026  VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA   1085

Qy    119  SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI   178
           ||:||||||||:  :||:|||||||||||||||||||||  :||:||||||||::||:
Db   1086  SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI   1143

Qy    179  NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG   238
           :||:|||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db   1144  DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG   1203

Qy    239  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF   298
           :||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204  SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF   1263

Qy    299  TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI   358
           ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db   1264  TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI   1323

Qy    359  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV   418
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV   1383

Qy    419  SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP   455
           ||:|||:||||||:|||||||||:||    |||||||:|
Db   1384  SPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP   1422
```

```
RESULT 11
Q51817
ID   Q51817;                                PRELIMINARY;         PRT;   1732 AA.
AC   Q51817;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Porphypain.
GN   PRTP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RC   STRAIN=W12;
RX   MEDLINE=96213011; PubMed=8631659;
RA   Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA   Progulske-Fox A., Lantz M.S.;
RT   "Analysis of the prtP gene encoding porphypain, a cysteine proteinase
RT   of Porphyromonas gingivalis.";
RL   J. Bacteriol. 178:2734-2741(1996).
DR   EMBL; U42210; AAB06565.1; -.
DR   PIR; T30836; T30836.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1732 AA;  187875 MW;  654271DBEF7BCAE4 CRC64;

Query Match             82.0%;  Score 2034.5;  DB 2;  Length 1732;
Best Local Similarity   84.3%;  Pred. No. 5.2e-124;
Matches  387;  Conservative  19;  Mismatches  44;  Indels  9;  Gaps  6;
```

```
Qy     60  ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA   118
           ||:::.  |||::||||||||||||||||||||||||||||||||||||||||||||||
Db   1026  VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA   1085

Qy    119  SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI   178
           ||:||||||||:  :||:|||||||||||||||||||||  :||:||||||||::||:
Db   1086  SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI   1143

Qy    179  NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG   238
           :||:|||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db   1144  DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG   1203

Qy    239  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF   298
           :||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204  SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF   1263

Qy    299  TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI   358
           ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db   1264  TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI   1323

Qy    359  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV   418
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV   1383

Qy    419  SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP   455
           ||:|||:||||||:|||||||||:||    |||||||:|
Db   1384  SPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP   1422
```

```
RESULT 12
O52050
ID   O52050;                                PRELIMINARY;         PRT;   1732 AA.
AC   O52050;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine specific cysteine protease.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RC   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=98298016; PubMed=9632563;
RA   Lewis J.P., Macrina F.L.;
RT   "IS195, an insertion sequence-like element associated with protease
RT   genes in Porphyromonas gingivalis.";
RL   Infect. Immun. 66:3035-3042(1998).
DR   EMBL; AF017059; AAC26523.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE   1732 AA;  187931 MW;  B2337463D5CB5EA5 CRC64;

Query Match             81.5%;  Score 2020.5;  DB 2;  Length 1732;
```

```
Best Local Similarity  83.9%;   Pred. No. 4.2e-123;
Matches 385;  Conservative 19;  Mismatches 46;  Indels  9;  Gaps  6;

Qy      1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAIC  59
            ||||||||||  || ||||||||||||||||||||||||||||  :
Db    969 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy     60 ASSASY-INFEGPQNPNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDA 118
            :||: :|||||||||||||||||||||||||||||||||||||||||||||||||
Db   1026 VYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDA 1085

Qy    119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
            |||| |||||| ||||  :|:|| ||||||||||||||||||||||||||||||||| |
Db   1086 SNFTNALLEETTTAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy    179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG 238
            :|:||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db   1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy    239 TNVVAGFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
            :|||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy    299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
            ||||||||||||||||||||||:|| |||||||||||||||||||||||||||||||||
Db   1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy    359 LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
            ||||||||||||||||||||||||||  |:|||||||  ||||||||||||||||||||
Db   1324 LLDDIQFTMGGSPTPTDYTYTVYRDGTNIKEGLTETTFEEYCVATGNHEYCVEVKYTAGV 1383

Qy    419 SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWNDAP 455
            |||:|:||  ||||:||||||||: |  ||  | |||||
Db   1384 SPKKCVDVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422
```

```
Db    311 LYENFENGPVPNGWLVIDADGDGFSW------------GHYLNAYDAFPGHNGGHCSLSASYVPG 363

Qy     69 EGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDASNFANALLEE 128
Db    364 IGPVTPDNYLIITPKVE--GAKRVKYWVSTQDANWAAEHYAVMASTTGTAVGDFV-ILFEE 420

Qy    129 VLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDV----- 183
Db    421 TMTAKPT-------------GAWYERTINLPEGTKYIAWRHYNCTDIYFLKLDDITVFGT 467

Qy    184 SNFANALLEEVLTAKTVVTAP---------------------------EIKANG------ 189
Db    468 PASEPEVPTDFVVSLIENNKGRLKWNYPNGYEPDKTDDKDPLQLAGYNIYANGSLLVHIQ 527

Qy    190 -------------KRADFTET------- 205
Db    528 DPTVLEYIDETYSSRDDQVEVEYCVTAVYNDNIESQSVCDKLIYDSQSDIILYEGFEAGS 587

Qy    206 APAEWTTIDADGDGQGWLCLSSGQLGW-LTAHGGTNVVASFSWNGM--ALNPDNYLISKD 262
Db    588 IPEGWLLIDADGDNVNW------DYYPWTMYGHDSEKCIASPSYLPMIGVLTPDNYLVTPR 642

Qy    263 VTGATKVKYYYAVNDG-FPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTE 321
Db    643 LEGAKLVKYWVSAQDAVYSAEHYAVMVSTTGTAVEDFVLLFEET--------------MTAK 690

Qy    322 ADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDI------YTYTVYRDGT--- 363
Db    691 ANGA------WYERTITLPAGTKYIAWRHYDCTDMFFLLDDITVYRSTETVPEVTDFVV 745

Qy    364 ------QFTMGGSPTPTD------------YTYTVYRDGT---KIKEGLTETTF 396
Db    746 SLIENNKGRLKWNYPNGYEPDKTDDKKPLQLTGYNIYANGSLLVHIQDPTVLEYIDETYS 805

Qy    397 EEDGVATGNHEYCVEVKYTAGV-SPKEC--VNVTV--DPVQ 432
Db    806 SRDGQV-EMEYCVTAVYNDNIESQSVCDKLNYTITSLDNIQ 845
```

RESULT 13
Q9F4J0
```
ID   Q9F4J0      PRELIMINARY;     PRT;   925 AA.
AC   Q9F4J0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative outer membrane protein PG57.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA   Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT   "Identification of vaccine candidates from genomic analysis of
RT   Porphyromonas gingivalis.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY007522; AAG24228.1; -.
DR   InterPro; IPR003961; FN_III.
DR   SMART; SM00060; FN3; 1.
SQ   SEQUENCE 925 AA; 103632 MW; 5FP219BD6914DAEB CRC64;

Query Match   27.5%;          DB 2;   Length 925;
Best Local Similarity 30.2%;   Pred. No. 4e-36;
Matches 176;  Conservative 62;  Mismatches 132;  Indels 212;  Gaps 22;

Qy    16 LSESFENG-IPASWKTIDADGDGNNWTTPPGG------TSFAGHNSAICASSASYINF 68
```

RESULT 14
Q9KIB3
```
ID   Q9KIB3      PRELIMINARY;     PRT;   312 AA.
AC   Q9KIB3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical outer membrane protein PG27.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C.;
RT   "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF237555; AAF81413.1; -.
SQ   SEQUENCE 312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;

Query Match   22.6%;   Score 561.5;   DB 2;   Length 312;
Best Local Similarity 32.8%;   Pred. No. 7e-29;
Matches 136;  Conservative 73;  Mismatches 23;  Indels 183;  Gaps 8;

Qy    16 LSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYINFEGPQNPD 75
```

```
Db                ||||||:||||| |||||||||||| ||   ||||   |:| :  | :||||||||:|||
          4 LSESFESGIPAIWKTIDADGDGYNWMHL-----TNFTG-QSGLCVSSASYIGGVGALTPD  57

Qy         76 NYLVTPELSLPNGG--TLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAK 133
              |||:||||||  :  ::||||||  :||||| | || :  |:||:|||||||||||||
Db         58 NYLITPELKLPTDALVEIIYWVCTQDLTAPSEHYAVYSSSTGNNAADFVNLLYEETLTAK 117

Qy        134 TVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGKRAD 193
              : :|    |::|| | : | | |||||  ||||||:|||||  |||||||||:|
Db        118 R--IQSPELIRGNRTQGVWYQRKVVLPNDTKYVAFRHFNSTDNFWLNLDEVSIL------- 169

Qy        194 FTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTNVVASFSWNGMALN 253

Db        170                                                              169

Qy        254 PDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGG 313
                                                               ||
Db        170 ----------------------------------------------YTP---------- 172

Qy        314 ARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTP 373
                                          ||                --
Db        173 ---------------------LP------RRAPC---------------PHP 182

Qy        374 TDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTV 428
              |||:||||||||:|| ||| || |   :   :::| ||||||||| ||:|:|
Db        183 GGYTYSVFRDGQKIASGLSALAYIDTDVPYGTQDYCVQNYLQGDSYKVCKNITVV 237


RESULT 15
Q9XBU9
ID   Q9XBU9        PRELIMINARY;      PRT;   293 AA.
AC   Q9XBU9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Immunoreactive 32 kDa antigen PG25.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF155351; AAD38410.1; -.
SQ   SEQUENCE  293 AA;  32272 MW;  CC03EAC241F7F6F1 CRC64;

Query Match             9.4%;  Score 233;  DB 2;  Length 293;
Best Local Similarity  59.0%;  Pred. No. 1.7e-07;
  Matches  46;  Conservative   7;  Mismatches  23;  Indels  2;  Gaps   1;

Qy        376 YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPVQFNP 435
              ||||||||||:|| ||| || |   :   :::| ||||||||| ||:|:|||:|  ||
Db         77 YTYTVYRDGVKIASGLTEPTFLDEDVPAGEHTYCVEVQYQGGVSDKVCVDVEVK--DFKP 134

Qy        436 VQNLTGSAVGQKVTLKWD 453
              | |||| :| :  ::||
Db        135 VTNLTGTASNDEVSLDWD 152
```

Search completed: May 18, 2004, 11:46:39
Job time : 34.3368 secs

```
          Copyright    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:  May 18, 2004, 11:38:40 ; Search time 14.0888 Seconds
                                   (without alignments)
                                   1670.936 Million cell updates/sec

Title:    US-08-570-311-20
Perfect score: 2480
Sequence: 1 GTPNPNPNPCTTTLSESF.........QNLTGSAVGQKVTLKMDAPN 456

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:  389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:    389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued Patents AA:*
    1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
    2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
    3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
    4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
    5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
    6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2480	100.0	456	2	US-08-570-311-18	Sequence 18, Appl
2	2480	100.0	456	2	US-08-570-311-20	Sequence 20, Appl
3	2480	100.0	2628	2	US-08-570-311-14	Sequence 14, Appl
4	2436	98.2	450	2	US-08-570-311-16	Sequence 16, Appl
5	2351	94.8	439	2	US-08-570-311-22	Sequence 22, Appl
6	2082	84.0	1477	4	US-09-482-500A-1	Sequence 1, Appl
7	2082	84.0	1687	2	US-08-570-311-29	Sequence 29, Appl
8	2082	84.0	1704	3	US-08-336-308A-10	Sequence 6, Appl
9	2082	84.0	1704	3	US-08-822-324-6	Sequence 6, Appl
10	2082	84.0	1704	3	US-09-490-931-10	Sequence 10, Appl
11	2070	83.5	1706	3	US-09-066-330-10	Sequence 10, Appl
12	2066	83.3	1087	2	US-09-570-311-8	Sequence 8, Appl
13	2066	83.3	1087	2	US-08-353-485-8	Sequence 8, Appl
14	2066	83.3	1358	2	US-08-570-311-27	Sequence 27, Appl
15	2039.5	82.2	1732	4	US-09-066-330-11	Sequence 11, Appl
16	2034.5	82.0	1732	2	US-08-570-311-10	Sequence 10, Appl
17	2034.5	82.0	1732	2	US-08-353-485-10	Sequence 10, Appl
18	546.5	22.0	497	2	US-08-570-311-2	Sequence 2, Appl
19	546.5	22.0	497	2	US-08-353-485-2	Sequence 2, Appl
20	296.5	12.0	942	1	US-08-141-324-14	Sequence 14, Appl
21	296.5	12.0	942	1	US-08-541-902-14	Sequence 14, Appl
22	236	9.5	49	3	US-08-822-324-18	Sequence 18, Appl
23	189	7.6	46	3	US-08-822-324-9	Sequence 9, Appl
24	137	5.5	1833	4	US-08-621-944A-4	Sequence 4, Appl
25	137	5.5	1833	4	US-08-945-567D-4	Sequence 4, Appl
26	137	5.5	1992	4	US-08-621-944A-3	Sequence 3, Appl
27	137	5.5	1992	4	US-08-945-567D-3	Sequence 3, Appl
28	134.5	5.4	2736	4	US-09-252-991A-30227	Sequence 30227, A
29	132.5	5.3	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
30	131.5	5.3	492	4	US-09-482-500A-2	Sequence 2, Appli
31	131.5	5.3	737	1	US-08-119-361-5	Sequence 5, Appli
32	131.5	5.3	737	3	US-08-336-308A-4	Sequence 4, Appli
33	131.5	5.3	737	3	US-08-822-324-4	Sequence 4, Appli
34	131.5	5.2	737	3	US-09-490-931-4	Sequence 10, Appl
35	131.5	5.2	2123	3	US-09-968-685A-10	Sequence 10, Appl
36	129.5	5.2	1638	4	US-09-071-035-258	Sequence 258, App
37	129.5	5.2	1638	4	US-09-071-035-262	Sequence 262, App
38	129.5	5.2	1638	4	US-09-071-035-266	Sequence 266, App
39	129.5	5.2	1747	4	US-09-134-000C-5999	Sequence 5999, Ap
40	129	5.2	24	4	US-09-066-330-6	Sequence 6, Appli
41	128	5.1	25	4	US-08-822-324-19	Sequence 19, Appl
42	126	5.1	2411	4	US-09-268-347-36	Sequence 36, Appl
43	123.5	5.0	669	4	US-09-071-035-264	Sequence 264, App
44	123.5	5.0	2048	4	US-09-268-347-48	Sequence 48, Appl
45	121	4.9	509	3	US-08-822-324-8	Sequence 8, Appli

ALIGNMENTS

```
RESULT 1
US-08-570-311-18
; Sequence 18, Application US/0857O311
; Patent No. 5824791
; GENERAL INFORMATION:
;  APPLICANT:  Progulske-Fox, Ann
;  APPLICANT:  Tumwasorn, Somying
;  APPLICANT:  Lepine, Guylaine
;  APPLICANT:  Han, Naiming
;  APPLICANT:  Lantz, Marilyn
;  APPLICANT:  Patti, Joseph
;  TITLE OF INVENTION:  Cloned Porphyromonas gingivalis Genes
;  TITLE OF INVENTION:  and Probes for the Detection of Periodontal Disease
;  NUMBER OF SEQUENCES: 29
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 18:
;  SEQUENCE CHARACTERISTICS:
```

```
;    LENGTH: 456 amino acids
;    TYPE: amino acid
;    TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-570-311-18

Query Match     100.0%;  Score 2480; DB 2; Length 456;
Best Local Similarity  100.0%; Pred. No. 2.8e-204;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQONPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQONPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

```
RESULT 2
US-08-570-311-20
; Sequence 20, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Ted W. Whitlock
;     STREET: 2421 N.W. 41st Street, Suite A-1
;     CITY: Gainesville
;     STATE: FL
;     COUNTRY: USA
;     ZIP: 32606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/570,311
;     FILING DATE:
;     CLASSIFICATION: 424
```

```
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/353,485
;     FILING DATE: 09-DEC-1994
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/647,119
;     FILING DATE: 25-JAN-1991
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/241,640
;     FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Whitlock, Ted W.
;     REGISTRATION NUMBER: 36,965
;     REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (904) 375-8100
;     TELEFAX:   (904) 372-5800
; INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 456 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-570-311-20

Query Match     100.0%;  Score 2480; DB 2; Length 456;
Best Local Similarity  100.0%; Pred. No. 2.8e-204;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQONPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQONPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||||||||||||||||||||||||||||||
Db  421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
```

```
RESULT 3
US-08-570-311-14
; Sequence 14, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
```

```
;,   APPLICANT: Patti, Joseph
;,   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;,   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;,   NUMBER OF SEQUENCES: 29
;,   CORRESPONDENCE ADDRESS:
;,    ADDRESSEE: Ted W. Whitlock
;,    STREET: 2421 N.W. 41st Street, Suite A-1
;,    CITY: Gainesville
;,    STATE: FL
;,    COUNTRY: USA
;,    ZIP: 32606
;,   COMPUTER READABLE FORM:
;,    MEDIUM TYPE: Floppy disk
;,    COMPUTER: IBM PC compatible
;,    OPERATING SYSTEM: PC-DOS/MS-DOS
;,    SOFTWARE: PatentIn Release #1.0, Version #1.25
;,   CURRENT APPLICATION DATA:
;,    APPLICATION NUMBER: US/08/570,311
;,    FILING DATE:
;,    CLASSIFICATION: 424
;,   PRIOR APPLICATION DATA:
;,    APPLICATION NUMBER: US 08/353,485
;,    FILING DATE: 09-DEC-1994
;,    CLASSIFICATION: 424
;,   PRIOR APPLICATION DATA:
;,    APPLICATION NUMBER: US 07/647,119
;,    FILING DATE: 25-JAN-1991
;,    CLASSIFICATION: 424
;,   PRIOR APPLICATION DATA:
;,    APPLICATION NUMBER: US 07/241,640
;,    FILING DATE: 08-SEP-1988
;,   ATTORNEY/AGENT INFORMATION:
;,    NAME: Whitlock, Ted W.
;,    REGISTRATION NUMBER: 36,965
;,    REFERENCE/DOCKET NUMBER: UF15.C3
;,   TELECOMMUNICATION INFORMATION:
;,    TELEPHONE: (904) 375-8100
;,    TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 14:
;,   SEQUENCE CHARACTERISTICS:
;,    LENGTH: 2628 amino acids
;,    TYPE: amino acid
;,    TOPOLOGY: linear
;,   MOLECULE TYPE: protein
US-08-570-311-14

Query Match       100.0%;  Score 2480; DB 2; Length 2628;
Best Local Similarity  100.0%;  Pred. No. 3.8e-203;
Matches 456; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy      1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    950 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 1009

Qy     61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1069

Qy    121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKVQLPAGTKYVAFRHFGCTDFFWINL 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKVQLPAGTKYVAFRHFGCTDFFWINL 1129

Qy    181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 1189

Qy    241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 1249

Qy    301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1309
```

```
Qy    361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 1369

Qy    421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
            ||||||||||||||||||||||||||||||||||||
Db   1370 KECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 1405


RESULT 4
US-08-570-311-16
; Sequence 16, Application US/08570311
; Patent No. 5824791
;, GENERAL INFORMATION:
;,    APPLICANT: Progulske-Fox, Ann
;,    APPLICANT: Tumwasorn, Somying
;,    APPLICANT: Lepine, Guylaine
;,    APPLICANT: Han, Naiming
;,    APPLICANT: Lantz, Marilyn
;,    APPLICANT: Patti, Joseph
;,    TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;,    TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;,    NUMBER OF SEQUENCES: 29
;,    CORRESPONDENCE ADDRESS:
;,     ADDRESSEE: Ted W. Whitlock
;,     STREET: 2421 N.W. 41st Street, Suite A-1
;,     CITY: Gainesville
;,     STATE: FL
;,     COUNTRY: USA
;,     ZIP: 32606
;,    COMPUTER READABLE FORM:
;,     MEDIUM TYPE: Floppy disk
;,     COMPUTER: IBM PC compatible
;,     OPERATING SYSTEM: PC-DOS/MS-DOS
;,     SOFTWARE: PatentIn Release #1.0, Version #1.25
;,    CURRENT APPLICATION DATA:
;,     APPLICATION NUMBER: US/08/570,311
;,     FILING DATE:
;,     CLASSIFICATION: 424
;,    PRIOR APPLICATION DATA:
;,     APPLICATION NUMBER: US 08/353,485
;,     FILING DATE: 09-DEC-1994
;,     CLASSIFICATION: 424
;,    PRIOR APPLICATION DATA:
;,     APPLICATION NUMBER: US 07/647,119
;,     FILING DATE: 25-JAN-1991
;,     CLASSIFICATION: 424
;,    PRIOR APPLICATION DATA:
;,     APPLICATION NUMBER: US 07/241,640
;,     FILING DATE: 08-SEP-1988
;,    ATTORNEY/AGENT INFORMATION:
;,     NAME: Whitlock, Ted W.
;,     REGISTRATION NUMBER: 36,965
;,     REFERENCE/DOCKET NUMBER: UF15.C3
;,    TELECOMMUNICATION INFORMATION:
;,     TELEPHONE: (904) 375-8100
;,     TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 16:
;,    SEQUENCE CHARACTERISTICS:
;,     LENGTH: 450 amino acids
;,     TYPE: amino acid
;,     TOPOLOGY: linear
;,    MOLECULE TYPE: protein
US-08-570-311-16

Query Match        98.2%;  Score 2436; DB 2; Length 450;
Best Local Similarity  99.8%;  Pred. No. 1.6e-200;
Matches 449; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy      7 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 66
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60
```

REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 439 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-22

Query Match　　94.8%;　Score 2351; DB 2; Length 439;
Best Local Similarity　98.2%;　Pred. No. 3e-193;
Matches 431;　Conservative　3;　Mismatches　5;　Indels　0;　Gaps　0;

Qy　　1　GTPNPNPNPNPGGTTLSESFENGIPASWKTIDADGDGNWWTTTPPPGGTSFAGHNSAICA　60
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　1　GTPNPNPNPNPGGTTLSESFENGIPASWKTIDADGDGNWWTTTPPPGGTSFAGHNSAICV　60

Qy　　61　SSASYINFEGPONPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN　120
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　61　SSASYINFEGPONPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN　120

Qy　　121　FANALLEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL　180
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　121　FANALLEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL　180

Qy　　181　DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN　240
　　　　　|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　181　DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN　240

Qy　　241　VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTV　300
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　241　VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTV　300

Qy　　301　VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL　360
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　301　VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL　360

Qy　　361　DDIQFTMGGSPTPTDYTTYTVYRDGTKIKEGLIETTFEEDGVATGNHEYCVEVKYTAGVSP　420
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　361　DDIQFTMGGSPTPTDYTTYTVYRDGTKIKEGLIETTFEEDGVATGNHEYCVEVKYTAGVSP　420

Qy　　421　KECVNVTVDPVQFNPVQNL　439
　　　　　||||||||:::||||||||
Db　　421　KVCVNVTINPTQFNPVQNL　439

RESULT 6
US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Imamura, Takahisa
; APPLICANT: Potempa, Jan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
; FILE REFERENCE: 235.00160101
; CURRENT APPLICATION NUMBER: US/09/482,500A
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/115,869
; PRIOR FILING DATE: 1999-01-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-482-500A-1

Query Match　　84.0%;　Score 2082;　DB 4;　Length 1477;

（Second column — RESULT 5 alignment and bibliographic data）

Qy　　67　NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL　126
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　61　NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL　120

Qy　　127　EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK　186
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　121　EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK　180

Qy　　187　ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVVASFS　246
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　181　ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS　240

Qy　　247　WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETP　306
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　241　WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETP　300

Qy　　307　NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT　366
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　301　NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT　360

Qy　　367　MGGSPTPTDYTTYTVYRDGTKIKEGLIETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV　426
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　361　MGGSPTPTDYTTYTVYRDGTKIKEGLIETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV　420

Qy　　427　TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN　456
　　　　　||||||||||||||||||||||||||||||
Db　　421　TVDPVQFNPVQNLTGSAVGQKVTLKWDAPN　450

RESULT 5
US-08-570-311-22
; Sequence 22, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.

Best Local Similarity 86.2%; Pred. No. 2.1e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

```
Qy    3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASS 62
          ||||||||||||||||||||||||||||||||| |:|:   ||  : :    ||
Db  725 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 781

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
            : :          ||:||  :  ||| ||||||||||||||||||||||||||||||
Db  782 ESFGLGGIGVLTPDNYLITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 841

Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          ||||| :|:||: | |:||||||  :|||| |||| |||||||||||||:  || |::||
Db  842 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 899

Qy  182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
          :||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db  900 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 959

Qy  242 VASFSWNGMALNPDNYLISKDVTGATKYKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  960 VASFSWNGMALNPDNYLISKDVTGATKYKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1019

Qy  302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          |||||||||||||||||||| |:|||||||||||||||||||||||||||||||||||||
Db 1020 FEETPNGINKGGARFGLSTRANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1079

Qy  362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1080 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1139

Qy  422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          |||||: :|.|||||:||    ||||  |||.|| :
Db 1140 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1174
```

RESULT 7
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991

;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 29:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1687 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-29

Query Match   84.0%; Score 2082; DB 2; Length 1687;
Best Local Similarity 86.2%; Pred. No. 2.5e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

```
Qy    3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASS 62
          ||||||||||||||||||||||||||||||||| |:|:   ||  : :    ||
Db  935 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 991

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
            : :          ||:||  :  ||| ||||||||||||||||||||||||||||||
Db  992 ESFGLGGIGVLTPDNYLITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWNLD 181
          ||||| :|:||: | |:||||||  :|||| |||| |||||||||||||:  || |::||
Db 1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy  182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
          :||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db 1110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1169

Qy  242 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1170 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1229

Qy  302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          |||||||||||||||||||| |:|||||||||||||||||||||||||||||||||||||
Db 1230 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1289

Qy  362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1290 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1349

Qy  422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          |||||: :|.|||||:||         ||||.|| :
Db 1350 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1384
```

RESULT 8
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
GENERAL INFORMATION:
;   APPLICANT: Travis, James
;   APPLICANT: Potempa, Jan S.
;   APPLICANT: Barr, Philip J.
;   APPLICANT: Pavloff, Nadine
;   TITLE OF INVENTION: Porphyromonas gingivalis
;   TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
;   STREET: 5370 Manhattan Circle, Suite 201
;   CITY: Boulder
;   STATE: Colorado

GENERAL INFORMATION:
  APPLICANT: Potempa, Jan S.
  APPLICANT: Travis, James
  APPLICANT: Genco, Caroline A.
  TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
  TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
  TITLE OF INVENTION: METHODS
  NUMBER OF SEQUENCES: 24
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
    STREET: 5370 Manhattan Circle, Suite 201
    CITY: Boulder
    STATE: CO
    COUNTRY: US
    ZIP: 80303
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/822,324
    FILING DATE: 21-MAR-1997
    CLASSIFICATION: 536
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 60/013,945
    FILING DATE: 22-MAR-1996
  ATTORNEY/AGENT INFORMATION:
    NAME: Ferber, Donna M.
    REGISTRATION NUMBER: 33,878
    REFERENCE/DOCKET NUMBER: 103-95 WO
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (303) 488-8080
    TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 6:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1704 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-822-324-6

```
Query Match     84.0%;  Score 2082;  DB 3;  Length 1704;
Best Local Similarity   86.2%;  Pred. No. 2.6e-169;
Matches  392;  Conservative  17;  Mismatches  40;  Indels   6;  Gaps    4;

Qy     3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
           |||||:||||||||||||||||||||||||||||  :|:||   |:|:
Db   952 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy    63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
           :: |||||  :||||:|||  |:  ||| ||||||||||||||||||||||
Db  1009 ESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
           |||||  :||:|  :|||| ||:|||:|||:||:||||||||||||||  :|::|:|:|
Db  1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
           :|||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
           |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db  1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1307 DIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366
```

  COUNTRY: US
  ZIP: 80303
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/336,308A
    FILING DATE: 08-NOV-1994
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/119,361
    FILING DATE: 10-SEP-1993
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/265,441
    FILING DATE: 24-JUN-1994
  ATTORNEY/AGENT INFORMATION:
    NAME: Ferber, Donna M.
    REGISTRATION NUMBER: 33,878
    REFERENCE/DOCKET NUMBER: 21-93C
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (303) 499-8080
    TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 10:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1704 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-336-308A-10

```
Query Match     84.0%;  Score 2082;  DB 3;  Length 1704;
Best Local Similarity   86.2%;  Pred. No. 2.6e-169;
Matches  392;  Conservative  17;  Mismatches  40;  Indels   6;  Gaps    4;

Qy     3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
           |||||:||||||||||||||||||||||||||||  :|:||   |:|:
Db   952 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy    63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
           :: |||||  :||||:|||  |:  ||| ||||||||||||||||||||||
Db  1009 ESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
           |||||  :||:|  :|||| ||:|||:|||:||:||||||||||||||  :|::|:|:|
Db  1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy   182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
           :|||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
           |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db  1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1307 DIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy   422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
           ||||| :: |||||| |  :: ||:: ||||:|:
Db  1367 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

RESULT 9
US-08-822-324-6
; Sequence 6, Application US/08822324
; Patent No. 6129917

```
Qy    422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          |||||:|| ||||:||  | |||:||:||
Db   1367 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

RESULT 10
US-09-490-931-10
; Sequence 10, Application US/09490931
; Patent No. 6274718
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/490,931
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/336,308
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-490-931-10

Query Match   84.0%;   Score 2082; DB 3; Length 1704;
Best Local Similarity 86.2%; Pred. No. 2.6e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;
```

```
Qy      3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASS 62
          |||||||||:|  ||||::|||||||||||||||||    |||||||||||||:|
Db    952 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy     63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          ||: :|||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db   1009 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy    122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          |||||||:|| :||| :|||:||:|||| :|||:||||||||||||||||||||||||||
Db   1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy    182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
          :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 11
US-09-066-330-10
; Sequence 10, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275
; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-10

```
Query Match      83.5%;   Score 2070;  DB 4;  Length 1706;
Best Local Similarity  85.7%;  Pred. No. 2.7e-168;
Matches  390; Conservative  18; Mismatches  41;  Indels  6;  Gaps  4;
```

```
Qy      3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASS 62
          |||||||||:|  ||||::|||||||||||||||||    |||||||||||||:|
Db    954 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy     63 ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          ||: :|||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

Qy    122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          |||||||:|| :||| :|||:||:|||| :|||:||||||||||||||||||||||||||
Db   1071 TNALLEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy    182 DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNV 241
          :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1129 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1188

Qy    242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          :||||||||||||||||:|||||||||||:||||||||||||||||:|||||||||||||
Db   1189 VSSFSWNGMALNPDNYLLSKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy    302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db   1249 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy    362 DIQFTMGGSPTPTDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy    242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          :||||||||||||||||:|||||||||||:||||||||||||||||:|||||||||||||
Db   1187 VASFSWNGMALNPDNYLLSKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy    302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy    362 DIQFTMGGSPTPTDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307 DIQFTMGGSPTPTDYTVTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy    422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          |||||:|| ||||:||  | |||:||:||
Db   1367 ECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1401
```

Db   390 YSESFGLGGIGVLTPDNYLIITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 449

Qy   120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN 179
          ||||||::||||:||| :||||||||| :|::||||||||||||||||||: ||:|:
Db   450 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 507

Qy   180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGT 239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   508 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGT 567

Qy   240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   568 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 627

Qy   300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||:|:||||
Db   628 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYIL 687

Qy   360 LDDIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   688 LDDIQFTMGGSPTPTDYTYTVRDGTKIKEGJTETTFEEDGVATGNHEYCVEVKYTAGVS 747

Qy   420 PKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          ||||||||:::||||||||| ||||:||||:||
Db   748 PKVCVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 784

RESULT 13
US-08-353-485-8
; Sequence 8, Application US/08353485
; Patent No. 5830710
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Ted W. Whitlock
;; STREET: 2421 N.W. 41st Street, Suite A-1
;; CITY: Gainesville
;; STATE: FL
;; COUNTRY: USA
;; ZIP: 32606
; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/353,485
;; FILING DATE: 09-DEC-1994
;; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/647,119
;; FILING DATE: 25-JAN-1991
;; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/241,640
;; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
;; NAME: Whitlock, Ted W.
;; REGISTRATION NUMBER: 36,965
;; REFERENCE/DOCKET NUMBER: UF15.C2
; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (904) 375-8100
;; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:

Db   1309 DIQFTMGGSPTPTDYTYTVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy   422 ECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          :||||||: ||||||||||||:|| ||||:|| :||
Db   1369 KCVNVTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 12
US-08-570-311-8
; Sequence 8, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Ted W. Whitlock
;; STREET: 2421 N.W. 41st Street, Suite A-1
;; CITY: Gainesville
;; STATE: FL
;; COUNTRY: USA
;; ZIP: 32606
; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/570,311
;; FILING DATE:
;; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/353,485
;; FILING DATE: 09-DEC-1994
;; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/647,119
;; FILING DATE: 25-JAN-1991
;; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/241,640
;; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
;; NAME: Whitlock, Ted W.
;; REGISTRATION NUMBER: 36,965
;; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (904) 375-8100
;; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1087 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-8

Query Match   83.3%; Score 2066; DB 2; Length 1087;
Best Local Similarity   85.3%; Pred. No. 3.1e-168;
Matches 390; Conservative 18; Mismatches 43; Indels 6; Gaps 4;

Qy   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNWTTTPPPGGTSFAGHNSAICA 60
          |||||||||||||||||||||||||||||||||||| :||| : |:||
Db   333 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 389

Qy   61 SSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
          |: :: | ||| | || ||||||| |||| || |||||||||||||||||||

```
;   LENGTH: 1087 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-353-485-8

Query Match    83.3%;   Score 2066;  DB 2;  Length 1087;
Best Local Similarity  85.3%;   Pred. No. 3.1e-168;
Matches 390;  Conservative  18;  Mismatches  43;  Indels  6;  Gaps  4;

Qy      1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA  60
          ||||||||||||||||||||||||||||||||||||||    :|: ||
Db    333 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV  389

Qy     61 SSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS  119
          ||   |   ||||:| ||||||||||||||||||||||||||||||||||||||||
Db    390 YSESFGLGGIGVLTPDNYLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS  449

Qy    120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN  179
          ||| ||||||||||  ::| ||||    ::|||  || |||||||||||| :|:
Db    450 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID  507

Qy    180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGT  239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    508 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGT  567

Qy    240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    568 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  627

Qy    300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL  359
          ||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||
Db    628 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYIL  687

Qy    360 LDDIQFTMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    688 LDDIQFTMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  747

Qy    420 PKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  456
          |||||:||||||||   ||| :||||:|
Db    748 PKVCVNVTINPTQFNPVKNLKAQPDGDGVVLKWEAPS  784
```

```
RESULT 14
US-08-570-311-27
; Sequence 27, Application US/08570311
Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Ted W. Whitlock
;     STREET: 2421 N.W. 41st Street, Suite A-1
;     CITY: Gainesville
;     STATE: FL
;     COUNTRY: USA
;     ZIP: 32606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/570,311
;     FILING DATE:
;     CLASSIFICATION: 424
```

```
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/353,485
;     FILING DATE: 09-DEC-1994
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/647,119
;     FILING DATE: 25-JAN-1991
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/241,640
;     FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Whitlock, Ted W.
;     REGISTRATION NUMBER: 36,965
;     REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (904) 375-8100
;     TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 27:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1358 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-27

Query Match    83.3%;   Score 2066;  DB 2;  Length 1358;
Best Local Similarity  85.3%;   Pred. No. 4.3e-168;
Matches 390;  Conservative  18;  Mismatches  43;  Indels  6;  Gaps  4;

Qy      1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA  60
          ||||||||||||||||||||||||||||||||||||||    :|: ||
Db    604 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV  660

Qy     61 SSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS  119
          ||   |   ||||:| ||||||||||||||||||||||||||||||||||||||||
Db    661 YSESFGLGGIGVLTPDNYLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS  720

Qy    120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN  179
          ||| ||||||||||  ::| ||||    ::|||  || |||||||||||| :|:
Db    721 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID  778

Qy    180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGT  239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    779 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGT  838

Qy    240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    839 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  898

Qy    300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL  359
          ||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||
Db    899 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYIL  958

Qy    360 LDDIQFTMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    959 LDDIQFTMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  1018

Qy    420 PKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN  456
          |||||:||||||||   ||| :||||:|
Db   1019 PKVCVNVTINPTQFNPVKNLKAQPDGDGVVLKWEAPS  1055
```

```
RESULT 15
US-09-066-330-11
; Sequence 11, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric C.
;   APPLICANT: Bhogal, Peter S.
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;   FILE REFERENCE: Reynolds
```

```
;  CURRENT APPLICATION NUMBER: US/09/066,330A
;  CURRENT FILING DATE: 1998-09-15
;  EARLIER APPLICATION NUMBER: PN 6275
;  EARLIER FILING DATE: 1995-10-30
;  EARLIER APPLICATION NUMBER: PCT/AU96/00673
;  EARLIER FILING DATE: 1996-10-30
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 11
;  LENGTH: 1732
;  TYPE: PRT
;  ORGANISM: Porphyromonas gingivalis
US-09-066-330-11
```

```
Query Match       82.2%;  Score 2039.5;  DB 4;  Length 1732;
Best Local Similarity  84.5%;  Pred. No. 1.2e-165;
Matches  388;  Conservative  18;  Mismatches  44;  Indels  9;  Gaps  6;
```

```
Qy      1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
            |||||||||||| ||||||||||||||||||||||||||          :  |:|
Db    969 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy     60 ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDA 118
            | :|: |  ||||||||||||| |||||||||||||||||||||||||||||||||
Db   1026 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDA 1085

Qy    119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
            ||| |||||| : |:||| |:  |   :|: |||||||||||||||||| |   |:
Db   1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy    179 NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGGQWLCLSSGQLGWLTAHGG 238
            :|:|||||||||||||||||||||||||||||||||||||||||| |||||| ||||||
Db   1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGGQWLCLSSGQLDWLTAHGG 1203

Qy    239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF 298
            :||: ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy    299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
            |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy    359 LLDDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 LLDDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy    419 SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 455
            || ||||||: |||||||||||  : |   |||||||:|
Db   1384 SPKKCVNVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422
```

```
Search completed: May 18, 2004, 11:49:16
Job time : 16.0888 secs
```

```
              GenCore version 5.1.6
           Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May 18, 2004, 11:42:50 ; Search time 35.3211 Seconds
                                       (without alignments)
                                       3592.387 Million cell updates/sec

Title:       US-08-570-311-20
Perfect score: 2480
Sequence:    1 GTPNPNPNPGTTTLSESF........QNLTGSAVGQKVTLKWDAPN 456

Scoring table: BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: ,   1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications AA:*
             1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
             2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
             3: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
             4: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
             5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
             6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
             7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
             8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
             9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
            10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
            11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
            12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
            13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
            14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
            15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
            16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
            17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
            18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2070	83.5	1706	14	US-10-229-066-10	Sequence 10, Appl
2	2039.5	82.2	1732	14	US-10-229-066-11	Sequence 11, Appl
3	726	29.3	419	15	US-10-174-695-3	Sequence 3, Appl
4	710.5	28.6	419	15	US-10-174-695-5	Sequence 5, Appli
5	385	15.5	231	15	US-10-174-695-6	Sequence 6, Appli
6	232.5	9.4	196	15	US-10-174-695-4	Sequence 4, Appli
7	153	6.2	29	15	US-10-387-977-18	Sequence 18, Appl
8	153	6.2	29	15	US-10-387-977-21	Sequence 21, Appl
9	153	6.2	29	15	US-10-387-977-23	Sequence 23, Appl
10	153	6.2	29	15	US-10-387-977-24	Sequence 24, Appl
11	153	6.2	29	15	US-10-387-977-26	Sequence 26, Appl
12	147	5.9	29	15	US-10-387-977-17	Sequence 17, Appl
13	147	5.9	872	12	US-10-282-122A-55467	Sequence 55467, A
14	145	5.8	29	15	US-10-387-977-25	Sequence 25, Appl
15	143	5.8	509	15	US-10-387-977-101	Sequence 101, App

16	137	5.5	1833	14	US-10-175-282-4	Sequence 4, Appli
17	137	5.5	1833	14	US-10-175-275-4	Sequence 4, Appli
18	137	5.5	1992	14	US-10-175-282-3	Sequence 3, Appli
19	137	5.5	1992	14	US-10-175-275-3	Sequence 3, Appli
20	136	5.5	29	15	US-10-387-977-20	Sequence 20, Appl
21	134	5.4	25	15	US-10-387-977-77	Sequence 77, Appl
22	134	5.4	25	15	US-10-387-977-80	Sequence 80, Appl
23	134	5.4	27	15	US-10-387-977-11	Sequence 11, Appl
24	134	5.4	27	15	US-10-387-977-12	Sequence 12, Appl
25	134	5.4	27	15	US-10-387-977-13	Sequence 13, Appl
26	132.5	5.3	555	14	US-10-156-761-14592	Sequence 14592, A
27	132.5	5.3	691	15	US-10-369-493-23473	Sequence 23473, A
28	131.5	5.3	2122	9	US-09-813-214A-9	Sequence 9, Appli
29	131.5	5.3	2468	12	US-10-282-122A-66335	Sequence 66335, A
30	131.5	5.3	2468	14	US-10-246-330-4	Sequence 4, Appli
31	131	5.3	960	15	US-10-311-879-15	Sequence 15, Appl
32	130	5.2	25	15	US-10-387-977-81	Sequence 81, Appl
33	130	5.2	1946	12	US-10-282-122A-62947	Sequence 62947, A
34	129.5	5.2	1638	12	US-10-206-576-258	Sequence 258, App
35	129.5	5.2	1638	12	US-10-206-576-262	Sequence 262, App
36	129.5	5.2	1638	12	US-10-206-576-266	Sequence 266, App
37	129.5	5.2	1728	12	US-10-282-122A-56997	Sequence 56997, A
38	129	5.2	24	14	US-10-229-066-6	Sequence 6, Appli
39	129	5.2	2435	12	US-10-282-122A-47453	Sequence 47453, A
40	127	5.1	25	15	US-10-387-977-83	Sequence 83, Appl
41	127	5.1	25	15	US-10-387-977-84	Sequence 84, Appl
42	127	5.1	25	15	US-10-387-977-85	Sequence 85, Appl
43	126	5.1	1649	15	US-10-369-493-18460	Sequence 18460, A
44	123.5	5.0	491	15	US-10-387-977-100	Sequence 100, App
45	123.5	5.0	669	12	US-10-206-576-264	Sequence 264, App

```
                              ALIGNMENTS

RESULT 1
US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
;  APPLICANT: Reynolds, Eric C.
;  APPLICANT: Bhogal, Peter S.
;  APPLICANT: Slakeski, Nada
;  TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;  FILE REFERENCE: Reynolds
;  CURRENT APPLICATION NUMBER: US/10/229,066
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: US/09/066,330
;  PRIOR FILING DATE: 1998-09-15
;  PRIOR APPLICATION NUMBER: PN 6275
;  PRIOR FILING DATE: 1995-10-30
;  PRIOR APPLICATION NUMBER: PCT/AU96/00673
;  PRIOR FILING DATE: 1996-10-30
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;  LENGTH: 1706
;  TYPE: PRT
;  ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

  Query Match             83.5%;  Score 2070;  DB 14;  Length 1706;
  Best Local Similarity   85.7%;  Pred. No. 3.9e-176;
  Matches  390;  Conservative   18;  Mismatches   41;  Indels   6;  Gaps    4;

Qy          3 PNPNPNPNPGTTTLSESFENGILVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 62
              PNPNPNPNPGTTTLSESFENGILVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF
Db        954 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASS 1010

Qy         63 ASY-INFEGPQNPDNYLVTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
              ||||::||||:|:|||||:|||||:||:||||||||||||||||||||||||||||||||
Db       1011 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070
```

```
Qy   122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          :|||||||| :|:|||:|||||||| |||| |||||||||||||| |:|:||
Db  1071  TNALLEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy   182  DVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWICLSSGQLGWLTAHGGTNV 241
          :|||||||||||||||||||||||||||||||||||||||||:||||
Db  1129  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWICLSSGQLDWLTAHGGTNV 1188

Qy   242  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          :|||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1189  VSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy   302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1249  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy   362  DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1309  DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy   422  ECVNVTVDPVQRNPVQNLTGSAVGQKVTLKWDAPN 456
          :|||||| :||||:|
Db  1369  KCVNVTVNSTQFNPVKNLKAQPDGGDVVLKWEAPS 1403
```

RESULT 2
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

Query Match   82.2%; Score 2039.5; DB 14; Length 1732;
Best Local Similarity 84.5%; Pred. No. 2.2e-173;
Matches 388; Conservative 18; Mismatches 44; Indels 9; Gaps 6;

```
Qy     1  GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
          ||||||||||||  ||||||||||||||||||||||||  ||:||
Db   969  GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGHGWKPGNAPG---IAGYNSNGC 1025

Qy    60  ASSASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
          |:||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  1026  VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy   119  SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
          |||:||||||| ||||||||||||||||||||||||| |||
Db  1086  SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy   179  NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWICLSSGQLGWLTAHGG 238
          :||: |||||||||||||||||||||||||||||||||||||||||:||||
Db  1144  DLDDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWICLSSSQLDWLTAHGG 1203

Qy   239  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
          ||||:|||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  1204  SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy   299  TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1264  TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy   359  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1324  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy   419  SPKECVNVTVDPVQFNPVQNLTGSAV--GQKVTLKWDAP 455
          |||:||||| |:||||||:
Db  1384  SPKKCVNVTVNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422
```

RESULT 3
US-10-174-695-3
; Sequence 3, Application US/10/174,695
; Publication No. US2003023022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 5292820000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

Query Match   29.3%; Score 726; DB 15; Length 419;
Best Local Similarity 75.3%; Pred. No. 2.3e-56;
Matches 143; Conservative 11; Mismatches 30; Indels 6; Gaps 4;

```
Qy     3  PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASS 62
          |||||||||||||||||||||||||||||| ||:||||
Db   235  PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGHGWKPGNAPG---IAGYNSNGCVYS 291

Qy    63  ASY-INFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          |: ||||||||||||||||||||||||||||||||||||||||||||||
Db   292  ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 351

Qy   122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          :|||||||| :|:|||:|||||||||| |:|:||
Db   352  TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 409

Qy   182  DVEIKANGKR 191
          :|||||||||
Db   410  EVEIKANGKR 419
```

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US2003023022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 5292820000700
; CURRENT APPLICATION NUMBER: US/10/174,695

```
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-5

Query Match        28.6%;  Score 710.5;  DB 15;  Length 419;
Best Local Similarity   73.6%;  Pred. No. 5.6e-55;
Matches  142;  Conservative  12;  Mismatches  32;  Indels  7;  Gaps  5;

Qy    1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAIC 59
                     |||||||||||||||||||||||||||||||:||    |||:  ::|
Db  232 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 288

Qy   60 ASSASY-INFEGPQONPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
           |||:     |||  ||||||||||||||||||||||||||||||||||||||||||
Db  289 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 348

Qy  119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
         |||:|||||| :||::|    :|||  :|||||:|||||:|||||||||||||  ::|:
Db  349 SNFTNALLEEITTAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 406

Qy  179 NLDDVEIKANGKR 191
         :|| |||||||||
Db  407 DLDEVEIKANGKR 419
```

RESULT 5
US-10-174-695-6
```
; Sequence 6, Application US/10174695
; Publication No. US2003023 2022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-6

Query Match        15.5%;  Score 385;  DB 15;  Length 231;
Best Local Similarity   41.4%;  Pred. No. 3.9e-26;
Matches  103;  Conservative  31;  Mismatches  69;  Indels  46;  Gaps  11;

Qy  213 IDADGDGQGWLCLSSGQLGWLTAHGGTNVV-ASFSWNGMALNPDNYLISKD--VTGATKV 269
         |||| :||  |||:   | :|||  |  |||     :|:||:|:||||||    |||||
Db    3 LDADHNTFGSVIPATGPL--FTGTASSNLYSANFEYL-IPANADPVVTTQNIIVTGCQEV 59

Qy  270 KYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVEETPNGINKGGARFGLSTEADGAKPQS 329
         |   :          ::||           :|||||||          :      ||:
Db   60 V------IPGGVY----------DYCITNPEPASG-------KMWIAGDGGNQPA 91

Qy  330 VWIERTVDLPAGTKYV-AFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIK 388
          :  :  ||||||||   ||                             ||||||||||
```

```
Db   92 RYDDFTFE--AGKKYIFTMRRAGMGDGTDMEVED------DSPASYTYTYRDGTKIK 141

Qy  389 EGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVD-PVQFNPVQNLTGSAVGQK 447
         |||| ||||||||| |||||||||||||||||||| ||| | || ||:||||||:||||
Db  142 EGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSSVGQK 201

Qy  448 VTLKWDAPN 456
         |||||||||
Db  202 VTLKWDAPN 210
```

RESULT 6
US-10-174-695-4
```
; Sequence 4, Application US/10174695
; Publication No. US2003023 2022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-4

Query Match          9.4%;  Score 232.5;  DB 15;  Length 196;
Best Local Similarity   39.1%;  Pred. No. 1.5e-12;
Matches   61;  Conservative  13;  Mismatches  63;  Indels  19;  Gaps  5;

Qy    3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICASS 62
         |||||||||||||||||||||||||||||||:||     |||:  ::|  ::|
Db   44 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVVL 100

Qy   63 ASYINFEGPQONPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFA 122
             :|::  ::     :|  :|     : :|  : ||  |  |||  :
Db  101 DNSAKIDRNQEINVY-NTAEYAKTNNAPI------KVVGYADE-----KTGTAAYNM- 145

Qy  123 NALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQL 158
          | :|::||  :  :::  :|  :   ::  :|: |:
Db  146 -KLSERRAKAVAKMLEKYGVSADRITIEWKGSSEQI 180
```

RESULT 7
US-10-387-977-18
```
; Sequence 18, Application US/10387977
; Publication No. US2004000 5276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
```

```
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 18
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-18

Query Match           6.2%;  Score 153;  DB 15;  Length 29;
Best Local Similarity    100.0%;  Pred. No. 1.3e-06;
Matches   29;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
              |||||||||||||||||||||||||||||
Db      1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 8
US-10-387-977-21
;   Sequence 21, Application US/10387977
;   Publication No. US2004005276A1
;   GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928200030l
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 21
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-21

Query Match           6.2%;  Score 153;  DB 15;  Length 29;
Best Local Similarity    100.0%;  Pred. No. 1.3e-06;
Matches   29;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
              |||||||||||||||||||||||||||||
Db      1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 9
US-10-387-977-23
; Sequence 23, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928200030l
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 23
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-23

Query Match           6.2%;  Score 153;  DB 15;  Length 29;
Best Local Similarity    100.0%;  Pred. No. 1.3e-06;
Matches   29;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
              |||||||||||||||||||||||||||||
Db      1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 10
US-10-387-977-24
; Sequence 24, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928200030l
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
;   PRIOR APPLICATION NUMBER: AU PO 6528
;   PRIOR FILING DATE: 1997-04-30
;   NUMBER OF SEQ ID NOS: 105
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 24
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-387-977-24

Query Match           6.2%;  Score 153;  DB 15;  Length 29;
Best Local Similarity    100.0%;  Pred. No. 1.3e-06;
Matches   29;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
              |||||||||||||||||||||||||||||
Db      1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 11
US-10-387-977-26
; Sequence 26, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: O'Brien-Simpson, Neil Martin
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;   TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;   FILE REFERENCE: 52928200030l
;   CURRENT APPLICATION NUMBER: US/10/387,977
;   CURRENT FILING DATE: 2003-07-18
;   PRIOR APPLICATION NUMBER: US 09/423,056
;   PRIOR FILING DATE: 2000-03-22
;   PRIOR APPLICATION NUMBER: PCT/AU98/00311
;   PRIOR FILING DATE: 1998-04-30
```

; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-26

```
Query Match             6.2%;     Score 153;  DB 15;  Length 29;
Best Local Similarity   100.0%;   Pred. No. 1.3e-06;
Matches 29;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy     375  DYTYTVYRDGTKIKEGLTETTFEEDGVAT  403
            |||||||||||||||||||||||||||||
Db       1  DYTYTVYRDGTKIKEGLTETTFEEDGVAT  29
```

RESULT 12
US-10-387-977-17
; Sequence 17, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200030301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-17

```
Query Match             5.9%;     Score 147;  DB 15;  Length 29;
Best Local Similarity   96.6%;    Pred. No. 4.3e-06;
Matches 28;  Conservative 0;  Mismatches 0;  Indels 1;  Gaps 0;

Qy     375  DYTYTVYRDGTKIKEGLTETTFEEDGVAT  403
            ||||||||||||||| |||||||||||||
Db       1  DYTYTVYRDGTKIKEGLTATTFEEDGVAT  29
```

RESULT 13
US-10-282-122A-55467
; Sequence 55467, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55467
; LENGTH: 872
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-55467

```
Query Match             5.9%;     Score 147;  DB 12;  Length 872;
Best Local Similarity   21.3%;    Pred. No. 0.00064;
Matches 112;  Conservative 56;  Mismatches 205;  Indels 154;  Gaps 23;

Qy      10  NPGTTLSESFENGIPASWKTIDADGDGNWTTTPPGGTSFAGHNSAICASSASYINFE  69
            ||    :  :      : :          :|:        ::   :    ::  :
Db     293  NTGTGTHNVTVDTGLPS--VAFNAISDDNVLNAVEKGQDLSVSGTSANLAEGTVVTLN  350

Qy      70  GPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASST---GNDASNFANALL  126
            |       : :| :|        |:    :::         : :      :
Db     351  G---KNYAATTAAD---GTWSLTVPAADLAGLGDHYTLSATATNGVGNSVSNTANLLV  402

Qy     127  EEVLTAKT------VVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI  178
              :            |||   : :      || :    :   |
Db     403  DTALPTVTINTVAGDNVINAAEVAAGQTISG---KVANAEAG------------441

Qy     179  NLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGG  238
               :::||   :    |        :: : :        :||||||     ::  ::    ::
Db     442  NTVTVTIGGN---SYTATVQSDLIWSVNVPESVLTALGNGD--LTVSATV--TNGHGN  492

Qy     239  T-----NVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGT  293
                    :    :|     :: :|| ::       ||    : :     ||      :::
Db     493  TGAGEREIVIDASLPGLRIDT---VAGDDV---------INSIEHGQNLIVTGSSDGL  538

Qy     294  NAGDFTVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGT--------342
            ||||         ||||      | ||                 |||
Db     539  AAG------TTLVTVNGKTYAASVLADGTWSAAIPAADVGALAAGTVVTAAGQSAA  590

Qy     343  -KYVAFRHYNCSDLNVILL-------DDI-------QFTMGGSPTPTD---------375
                ||     : :|| :         ||:        | |:   | |   |
Db     591  GNPVTISHDVTVDLSAVAISIDAIATDDVINAAEKGADLVLSGSTSNVEEENQTVTITSGG  650

Qy     376  --YTYTVYRDG-------TKIKEG---LTETTFEEDGVATGNHEYCVEVKYTAGV  418
                 :|||            :|::|       :|||            :  :  |
Db     651  KTYTAKVDADGNWTATVPSADLAGLKDGDASVQVSVTNAHGNSASAGREYSVD-----AT  705

Qy     419  SPKECVNVTVDPVQFNPVQNLTGSA----------VGQKVTLKWD  453
                   :| |   |  |   |:   | :          |   |:| |
Db     706  AP----TVTIDTVAGDNVINASEAAGVAISGTTAEVGQTVTVTLD  748
```

```
Qy   112  SSTGNDASNFANALL--EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQ--LPAGTKYVA 166
          :|: .|| :|:     .| | |:  |:                      :| |:
Db   123  ASSPELTNIIDKVLMYEKATMPDKSYLEKVLLIAG--ADYSWNSQVGQPTIKYGMQYYy 180

Qy   167  FRHFGCTDFFWINLDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWL--C 224
          :|:.|:         | |:              :| | |
Db   181  NQEHGYTDVY-------------NYLKAPYTGCYSHLNTGVSFANYT---AHGSETAWADPL 226

Qy   225  LSSGQLGWLT------AHGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVND 277
          |::.||                  :|
Db   227  LTTSQLKALTNKDKYFLAIGNCCITAQFDY----VQP----CFGEVITRVKEKGAYAYIG 278

Qy   278  GFPGDH------------YAVMISKTGTNAGDFTVVF-EETPNGINKGGARFGLSTEA 322
          :| | |            : |: |
Db   279  SSPNSTWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSYNTVN------------ 326

Qy   323  DGAKPQSVWIERTVDLPAGT---KVVAFRHYNCSDLNYILLDDIQFTMGGSPTPDYT- 377
          :| |                  :| : :
Db   327  ----SIMWAGNLAATHAGNIGNITHIGAHY--WEAYHVLGDGSVMPYRAMPKTNTYTL 379

Qy   378  -------------YTVYRDGTKIKEGLTETTFEEDGVAT-----------GNHE 407
                         |||:|
Db   380  PASLPQNQASYSIQASAGSYVAISKDGVLYGTGVANAS----GVATVSMTKQITENGNYD 435

Qy   408  YCVEVKYTAGVSPKECVNVTVDPVQFNPVQNLTGSAVGQKVTLKWDAPN 456
          :| |                || ||||:|
Db   436  VVITRSNYLPVIKQIQVG---EPSPYQPVSNLTATTQGQKVTLKWEAPS 481
```

Search completed: May 18, 2004, 11:52:21
Job time : 36.3211 secs

RESULT 14
US-10-387-977-25
; Sequence 25, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-25

Query Match   5.8%;   Score 145; DB 15; Length 29;
Best Local Similarity   96.6%;   Pred. No. 6.6e-06;
Matches 28; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy   375  DYTYTVRDGTKIKEGLTETTFEEDGVAT 403
          |||||||||||| |||||||||||||||
Db     1  DYTYTVRDGTKIKWGLTETTFEEDGVAT 29
```

RESULT 15
US-10-387-977-101
; Sequence 101, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-101

Query Match   5.8%;   Score 143; DB 15; Length 509;
Best Local Similarity   20.0%;   Pred. No. 0.00066;
Matches 94; Conservative 66; Mismatches 189; Indels 120; Gaps 21;

```
Qy   54  HNSAICASSAS-YINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAV-YA 111
         :: .||| ||: ..                                 :  :  :
Db   67  YNDGLAASAAPVFLAIVG---DTDVISGEKGKKTKKVTDLYYSAVDGDYFPEMYTFRMS 122
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May 18, 2004, 11:33:39 ; Search time 45.2755 Seconds
                              (without alignments)
                              2739.638 Million cell updates/sec

Title:       US-08-570-311-22
Perfect score: 2388
Sequence:    1 GTPNPNPNPGTTLSESF........PKVCVNVTINPTQFNPVQNL 439

Scoring table: BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:        1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     A_Geneseq_29Jan04:*
               1: geneseqp1980s:*
               2: geneseqp1990s:*
               3: geneseqp2000s:*
               4: geneseqp2001s:*
               5: geneseqp2002s:*
               6: geneseqp2003as:*
               7: geneseqp2003bs:*
               8: geneseqp2004s:*

                          SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result                  *  Query
No.   Score   Match  Length  DB  ID          Description
----  -----   -----  ------  --  --          -----------
  1   2388   100.0   439    2  AAR96024   Aar96024 P. gingiv
  2   2388   100.0   439    2  AAW69492   Aaw69492 Haemagglu
  3   2388   100.0   2628   2  AAR96030   Aar96030 P. gingiv
  4   2388   100.0   2628   2  AAW69488   Aaw69488 Haemagglu
  5   2351    98.5   456    2  AAR96023   Aar96023 P. gingiv
  6   2351    98.5   456    2  AAR96022   Aar96022 P. gingiv
  7   2351    98.5   456    2  AAW69491   Aaw69491 Haemagglu
  8   2351    98.5   456    2  AAW69490   Aaw69490 Haemagglu
  9   2321    97.2   450    2  AAR96021   Aar96021 P. gingiv
 10   2321    97.2   450    2  AAW69489   Aaw69489 Haemagglu
 11   2058    86.2   1687   2  AAR96033   Aar96033 P. gingiv
 12   2058    86.2   1687   2  AAW69495   Aaw69495 Haemagglu
 13   2058    86.2   1687   2  AAR70188   Aar70188 Arg-gingi
 14   2058    86.2   1704   2  AAW34843   Aaw34843 Arg-gingi
 15   2058    86.2   1704   2  AAY67396   Aay67396 Arg-gingi
 16   2058    86.2   1704   4  AAU08938   Aau08938 P. gingiv
 17   2055    86.1   1087   2  AAR96028   Aar96028 P. gingiv
 18   2055    86.1   1087   2  AAW69486   Aaw69486 Haemagglu
 19   2055    86.1   1358   2  AAR96032   Aar96032 P. gingiv
 20   2055    86.1   1358   2  AAW69494   Aaw69494 Haemagglu
 21   2048    85.8   1706   2  AAW24786   Aaw24786 PrtR anti
 22   2025.5  84.8   1732   2  AAR96029   Aar96029 P. gingiv
 23   2025.5  84.8   1732   2  AAW24787   Aaw24787 PrtK anti
 24   2025.5  84.8   1732   2  AAW69487   Aaw69487 Haemagglu
 25   727     30.4   135    6  ABP55081   Abp55081 Porphyrom
```

```
 26   727    30.4            419  4  AAU03572  AAU03572  P. gingiv
 27   711.5  29.8            419  4  AAU03574  AAU03574  P. gingiv
 28   711    29.8            134  4  AAB49217  AAB49217  Peptide u
 29   691.5  29.0            921  2  AAY34522  AAY34522  Porphorym
 30   691.5  29.0            922  2  AAY34521  AAY34521  Porphorym
 31   691.5  29.0            925  2  AAY34520  AAY34520  Porphorym
 32   691.5  24.1            938  2  AAY34392  AAY34392  Porphorym
 33   575.5  23.9            377  2  AAY34359  AAY34359  Porphorym
 34   571.5  19.2            312  2  AAY34484  AAY34484  Porphorym
 35   458.5  19.2            497  2  AAR96025  AAR96025  P. gingiv
 36   458.5  13.6            497  2  AAW69483  AAW69483  Haemagglu
 37   325    12.7            970  2  AAR72458  AAR72458  Porphyrom
 38   303    10.6            231  4  AAU03575  AAU03575  P. gingiv
 39   252.5  10.1            991  2  AAR77313  AAR77313  Porphyrom
 40   242     9.8            49   2  AAW34805  AAW34805  Arg-speci
 41   233.5   8.3            196  4  AAU03573  AAU03573  P. gingiv
 42   199     8.3            293  2  AAY34483  AAY34483  Porphorym
 43   199     8.2            299  2  AAY34358  AAY34358  Porphorym
 44   195     6.4            46   2  AAW34798  AAW34798  Arg-speci
 45   153                    29   2  AAW83085  AAW83085  Peptide f
```

ALIGNMENTS

```
RESULT 1
AAR96024
ID   AAR96024 standard; protein; 439 AA.
XX
AC   AAR96024;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep4 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep4.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;       95WO-US016108.
XX
PR   09-DEC-1994;       94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30648.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 114-115; 153pp; English.
XX
CC   HArep4 (AAR96024) is the product of the HArep4 repeat unit (AAT30648) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see AAR96030). HArep4 and other hagA repeat
CC   unit products (see also AAR96021-23) can be obtd. from transformed host
CC   cells and used as vaccines to protect humans or animals against
CC   periodontal disease. Expression in Salmonella cells allows prodn. of live
CC   vaccine. HArep1-4 can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 439 AA;
```

XX
CC This sequence is encoded by a Porphyromonas gingivalis gene of the
CC invention. This sequence represents the hagA haemagglutinin protein. The
CC polypeptides are used to produce antibodies to organisms associated with
CC periodontal disease. The antibodies are also used in purification and
CC identification procedures. The genes and polypeptides are used as
CC vaccines against periodontal disease
XX
SQ Sequence 439 AA;

Query Match   100.0%;   Score 2388; DB 2; Length 439;
Best Local Similarity 100.0%;   Pred. No. 1.2e-185;
Matches 439;   Conservative 0;   Mismatches 0;   Indels 0; Gaps 0;

Qy   1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICV 60

Qy   61 SSASYINFEGPQNPDNVLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNVLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy   181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYRHYNCSDLNYILL 360

Qy   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy   421 KVCVNVTINPTQFNPVQNL 439
          |||||||||||||||||||
Db   421 KVCVNVTINPTQFNPVQNL 439

RESULT 3
AAR96030
ID AAR96030 standard; protein; 2628 AA.
XX
AC AAR96030;
XX
DT 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)
XX
DE P. gingivalis haemagglutinin hagA.
XX
KW Haemagglutinin; hagA; periodontal disease; vaccine; antibody.
XX
OS Porphyromonas gingivalis; strain 381.
XX
FH Key   Location/Qualifiers
FT Peptide   5..21
FT   /label= Sig_peptide
XX
PN WO961793G-A2.
XX
PD 13-JUN-1996.
XX
PF 11-DEC-1995; 95WO-US016108.
XX
PR 09-DEC-1994; 94US-00353485.

Query Match   100.0%;   Score 2388; DB 2; Length 439;
Best Local Similarity 100.0%;   Pred. No. 1.2e-185;
Matches 439;   Conservative 0;   Mismatches 0;   Indels 0; Gaps 0;

Qy   1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICV 60

Qy   61 SSASYINFEGPQNPDNVLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNVLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy   181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYRHYNCSDLNYILL 360

Qy   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy   421 KVCVNVTINPTQFNPVQNL 439
          |||||||||||||||||||
Db   421 KVCVNVTINPTQFNPVQNL 439

RESULT 2
AAW69492
ID AAW69492 standard; protein; 439 AA.
XX
AC AAW69492;
XX
DT 22-DEC-1998 (first entry)
XX
DE Haemagglutinin protein hagA, HArep4.
XX
KW Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS Porphyromonas gingivalis.
XX
PN US5824791-A.
XX
PD 20-OCT-1998.
XX
PF 11-DEC-1995; 95US-00570311.
XX
PR 08-SEP-1988; 88US-00241640.
PR 25-JAN-1991; 91US-00647119.
PR 09-DEC-1994; 94US-00353485.
XX
PA (UYFL) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR WPI; 1998-582627/49.
DR N-PSDB; AAV58879.
XX
PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT protease poly:peptide(s)).
XX
PS Claim 1; Col 139-144; 101pp; English.

XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30654.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 6; Page 93-101; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagA (AAR96030) was identified as the
CC   product of the hagA gene (AAT30654) isolated as an EcoRV fragment of
CC   genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
CC   24), can be obtd. from transformed host cells and used as a vaccine to
CC   protect humans or animals against periodontal disease. Expression in
CC   Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
CC   also be used to detect the presence of anti-P. gingivalis antibodies and
CC   to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-
CC   2003 to standardise OS field)
XX
SQ   Sequence 2628 AA;

     Query Match   100.0%; Score 2388; DB 2; Length 2628;
     Best Local Similarity   100.0%; Pred. No. 1.5e-184;
     Matches 439; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1862 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 1921

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1922 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1981

Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKVVAFRHFGCTDFFWINL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1982 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKVVAFRHFGCTDFFWINL 2041

Qy   181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2042 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 2101

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2102 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 2161

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2162 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 2221

Qy   361 DDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2222 DDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 2281

Qy   421 KVCVNVTINPTQFNPVQNL 439
          |||||||||||||||||||
Db   2282 KVCVNVTINPTQFNPVQNL 2300

RESULT 4
AAW69488
ID   AAW69488 standard; protein; 2628 AA.
XX
AC   AAW69488;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagA.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995; 95US-00570311.
XX
PR   08-SEP-1988; 88US-00241640.
PR   25-JAN-1991; 91US-00647119.
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58875.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 91-110; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 2628 AA;

     Query Match   100.0%; Score 2388; DB 2; Length 2628;
     Best Local Similarity   100.0%; Pred. No. 1.5e-184;
     Matches 439; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1862 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 1921

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1922 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1981

Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKVVAFRHFGCTDFFWINL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1982 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKVVAFRHFGCTDFFWINL 2041

Qy   181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2042 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 2101

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2102 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 2161

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2162 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 2221

Qy   361 DDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2222 DDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 2281

Qy   421 KVCVNVTINPTQFNPVQNL 439
          |||||||||||||||||||
Db   2282 KVCVNVTINPTQFNPVQNL 2300

RESULT 5
AAR96023
ID   AAR96023 standard; protein; 456 AA.
XX
AC   AAR96023;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep3 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep3.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tunwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30647.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 110-112; 153pp; English.
XX
CC   HArep3 (AAR96023) is the product of the HArep3 repeat unit (AAT30647) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep3 and other hagA repeat
CC   unit products (see also AAR96021-22 and AAR96024) can be obtd. from
CC   transformed host cells and used as vaccines to protect humans or animals
CC   against periodontal disease. Expression in Salmonella cells allows prodn.
CC   of live vaccine. HArep1-4 can also be used to detect the presence of anti
CC   -P. gingivalis antibodies and to raise monoclonal antibodies for
CC   diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 456 AA;

Query Match   98.5%;   Score 2351; DB 2; Length 456;
Best Local Similarity 98.2%;   Pred. No. 1.3e-182;
Matches 431;   Conservative 3;   Mismatches 5;   Indels 0;   Gaps 0;

Qy   1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
Qy   61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
         |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
Qy   121 FANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 FANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
Qy   181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTN 240
         |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTN 240
Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
Qy   361 DDIQFTMGGSPTPTDYTTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 DDIQFTMGGSPTPTDYTTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
Qy   421 KVCVNVTINPTQFNPVQNL 439
         ||||||| | |:|||||||
Db   421 KECVNVTVDPVQFNPVQNL 439

RESULT 6
AAR96022
ID   AAR96022 standard; protein; 456 AA.
XX
AC   AAR96022;
XX
DT   16-OCT-2003 (revised)
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep2 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep2.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tunwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30646.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 107-108; 153pp; English.
XX
CC   HArep2 (AAR96022) is the product of the HArep2 repeat unit (AAT30646) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep2 and other hagA repeat
CC   unit products (see also AAR96021 and AAR96023-24) can be obtd. from
CC   transformed host cells and used as vaccines to protect humans or animals
CC   against periodontal disease. Expression in Salmonella cells allows prodn.
CC   of live vaccine. HArep1-4 can also be used to detect the presence of anti
CC   -P. gingivalis antibodies and to raise monoclonal antibodies for
CC   diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 456 AA;

Query Match   98.5%;   Score 2351; DB 2; Length 456;
Best Local Similarity 98.2%;   Pred. No. 1.3e-182;
Matches 431;   Conservative 3;   Mismatches 5;   Indels 0;   Gaps 0;

Qy   1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
Qy   61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
         |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
Qy   121 FANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

```
Qy     1  GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICA    60

Qy    61  SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN   120
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    61  SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN   120

Qy   121  FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL   180

Qy   181  DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTN   240
          :|:||||||||||||||||||||||||||||||||||||||:||||||||:|||||||
Db   181  DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTN   240

Qy   241  VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV   300

Qy   301  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL   360

Qy   361  DDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  DDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP   420

Qy   421  KVCVNVTINPTQFNPVQNL   439
          ||||| |:|| |||||||||
Db   421  KECVNVTVDPVQFNPVQNL   439
```

```
RESULT 7
AAW69491
ID    AAW69491 standard; protein; 456 AA.
XX
AC    AAW69491;
XX
DT    22-DEC-1998 (first entry)
XX
DE    Haemagglutinin protein hagA, HArep3.
XX
KW    Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS    Porphyromonas gingivalis.
XX
PN    US5824791-A.
XX
PD    20-OCT-1998.
XX
PF    11-DEC-1995;      95US-00570311.
XX
PR    08-SEP-1988;      88US-00241640.
PR    25-JAN-1991;      91US-00647119.
PR    09-DEC-1994;      94US-00353485.
XX
PA    (UYFL ) UNIV FLORIDA.
PA    (UABR-) UAB RES FOUND.
XX
PI    Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR    WPI; 1998-582627/49.
DR    N-PSDB; AAV58878.
XX
PT    Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT    protease poly:peptide(s)).
XX
PS    Claim 1; Col 133-138; 101pp; English.
XX
CC    This sequence is encoded by a Porphyromonas gingivalis gene of the
CC    invention. This sequence represents the hagA haemagglutinin protein. The
CC    polypeptides are used to produce antibodies to organisms associated with
CC    periodontal disease. The antibodies are also used in purification and
CC    identification procedures. The genes and polypeptides are used as
CC    vaccines against periodontal disease
XX
SQ    Sequence 456 AA;

Query Match    98.5%;   Score 2351;  DB 2;  Length 456;
Best Local Similarity   98.2%;   Pred. No. 1.3e-182;
Matches 431;  Conservative 3;  Mismatches 5;  Indels 0;  Gaps 0;
```

```
Db   121  FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   181  DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTN   240
          :|:||||||||||||||||||||||||||||||||||||||:||||||||:|||||||
Db   181  DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLGWLTAHGGTN   240

Qy   241  VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV   300

Qy   301  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL   360

Qy   361  DDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  DDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP   420

Qy   421  KVCVNVTINPTQFNPVQNL   439
          ||||| |:|| |||||||||
Db   421  KECVNVTVDPVQFNPVQNL   439
```

```
RESULT 8
AAW69490
ID    AAW69490 standard; protein; 456 AA.
XX
AC    AAW69490;
XX
DT    22-DEC-1998 (first entry)
XX
DE    Haemagglutinin protein hagA, HArep2.
XX
KW    Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS    Porphyromonas gingivalis.
XX
PN    US5824791-A.
XX
PD    20-OCT-1998.
XX
PF    11-DEC-1995;      95US-00570311.
XX
PR    08-SEP-1988;      88US-00241640.
PR    25-JAN-1991;      91US-00647119.
PR    09-DEC-1994;      94US-00353485.
XX
PA    (UYFL ) UNIV FLORIDA.
PA    (UABR-) UAB RES FOUND.
XX
PI    Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR    WPI; 1998-582627/49.
DR    N-PSDB; AAV58877.
XX
PT    Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT    protease poly:peptide(s)).
XX
PS    Claim 1; Col 127-132; 101pp; English.
XX
CC    This sequence is encoded by a Porphyromonas gingivalis gene of the
CC    invention. This sequence represents the hagA haemagglutinin protein. The
```

```
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 456 AA;

Query Match         98.5%;  Score 2351;  DB 2;  Length 456;
Best Local Similarity  98.2%;  Pred. No. 1.3e-182;
Matches  431;  Conservative  3;  Mismatches  5;  Indels  0;  Gaps  0;

Qy    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPGGTSFAGHNSAICV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPGGTSFAGHNSAICA 60

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy  181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy  421 KVCVNVTINPTQFNPVQNL 439
          | |||||||::||||||||
Db  421 KECVNVTVDPVQFNPVQNL 439

RESULT 9
AAR96021
ID   AAR96021 standard; protein; 450 AA.
XX
AC   AAR96021;
XX
DT   16-OCT-2003  (revised)
DT   04-SEP-1996  (first entry)
XX
DE   P. gingivalis haemagglutinin hagA HArep1 product.
XX
KW   Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HArep1.
XX
OS   Porphyromonas gingivalis; strain 381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995; 95WO-US016108.
XX
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
```

```
DR   N-PSDB; AAT30645.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 4; Page 103-104; 153pp; English.
XX
CC   HArep1 (AAR96021) is the product of the HArep1 repeat unit (AAT30645) of
CC   the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
CC   haemagglutinin hagA (see also AAR96030). HArep1 and other hagA repeat
CC   unit products (see also AAR96022-24) can be obtd. from transformed host
CC   cells and used as vaccines to protect humans or animals against
CC   periodontal disease. Expression in Salmonella cells allows prodn. of live
CC   vaccine. HArep1-4 can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 450 AA;

Query Match         97.2%;  Score 2321;  DB 2;  Length 450;
Best Local Similarity  98.4%;  Pred. No. 3.4e-180;
Matches  426;  Conservative  3;  Mismatches  4;  Indels  0;  Gaps  0;

Qy    7 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPGGTSFAGHNSAICVSSASYI 66
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    1 PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPGGTSFAGHNSAICASSASYI 60

Qy   67 NFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120

Qy  127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIK 186
          |||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db  121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180

Qy  187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 246
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240

Qy  247 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300

Qy  307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360

Qy  367 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNV 426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV 420

Qy  427 TINPTQFNPVQNL 439
          :::||||||||
Db  421 TVDPVQFNPVQNL 433

RESULT 10
AAW69489
ID   AAW69489 standard; protein; 450 AA.
XX
AC   AAW69489;
XX
DT   22-DEC-1998  (first entry)
XX
DE   Haemagglutinin protein hagA, HArep1.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagA.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
```

```
XX
PF   11-DEC-1995;   95US-00570311.
XX
PR   08-SEP-1988;   88US-00241640.
PR   25-JAN-1991;   91US-00647119.
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58876.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 121-126; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagA haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 450 AA;

Query Match             97.2%;  Score 2321;  DB 2;  Length 450;
Best Local Similarity   98.4%;  Pred. No. 3.4e-180;
Matches  426;   Conservative   3;   Mismatches   4;   Indels   0;   Gaps   0;

Qy          7 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSSASYI 66
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICASSASYI 60

Qy         67 NFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120

Qy        127 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIK 186
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db        121 EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180

Qy        187 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 246
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS 240

Qy        247 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 306
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP 300

Qy        307 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360

Qy        367 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNV 426
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKBCVNV 420

Qy        427 TINPTQFNPVQNL 439
              ||::||||||||||
Db        421 TVDPVQFNPVQNL 433
```

```
DT   16-OCT-2003 (revised)

RESULT 11
AAR96033
ID   AAR96033 standard; protein; 1687 AA.
XX
AC   AAR96033;
XX
DT   04-SEP-1996 (first entry)
XX
DE   P. gingivalis haemagglutinin hagE.
XX
KW   Haemagglutinin; hagE; periodontal disease; vaccine; antibody.
XX
OS   Porphyromonas gingivalis; strain FDC381.
XX
PN   WO9617936-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995;   95WO-US016108.
XX
PR   09-DEC-1994;   94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR   WPI; 1996-287181/29.
DR   N-PSDB; AAT30656.
XX
PT   Porphyromonas gingivalis genes and proteins - used in the detection and
PT   vaccination against periodontal disease.
XX
PS   Claim 5; Page 138-143; 153pp; English.
XX
CC   P. gingivalis 381 haemagglutinin hagE (AAR96033) was identified as the
CC   product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
CC   The haemagglutinin can be obtd. from transformed host cells and used as a
CC   vaccine to protect humans or animals against periodontal disease.
CC   Expression in Salmonella cells allows prodn. of a live vaccine. The
CC   haemagglutinin can also be used to detect the presence of anti-P.
CC   gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC   appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1687 AA;

Query Match             86.2%;  Score 2058;  DB 2;  Length 1687;
Best Local Similarity   88.8%;  Pred. No. 5.8e-158;
Matches  389;   Conservative  12;   Mismatches  31;   Indels  6;   Gaps  4;

Qy          3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS 62
              |||||||||||||||||||||||||||||||||||||||||||| :|||||||||
Db        935 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 991

Qy         63 ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
              ::| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        992 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy        122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
              | : :||| |::|||||||||   :|:|::|||||||||||||||||||||:|:||
Db       1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy        182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 241
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1169

Qy        242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1229

Qy        302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1289

Qy        362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1349

Qy        422 VCVNVTINPTQFNPVQNL 439
```

```
Db   1350 ECVNVTINPTQFNPVKNL 1367
          ||||||||||||||:||

RESULT 12
AAW69495
ID   AAW69495 standard; protein; 1687 AA.
XX
AC   AAW69495;
XX
DT   22-DEC-1998 (first entry)
XX
DE   Haemagglutinin protein hagE.
XX
KW   Haemagglutinin protein; periodontal disease; vaccine; hagE.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5824791-A.
XX
PD   20-OCT-1998.
XX
PF   11-DEC-1995; 95US-00570311.
XX
PR   08-SEP-1988; 88US-00241640.
PR   25-JAN-1991; 91US-00647119.
PR   09-DEC-1994; 94US-00353485.
XX
PA   (UYFL ) UNIV FLORIDA.
PA   (UABR-) UAB RES FOUND.
XX
PI   Patti JM, Han N, Lantz M, Tumwasorn S, Proguleske-Fox A, Lepine G;
XX
DR   WPI; 1998-582627/49.
DR   N-PSDB; AAV58881.
XX
PT   Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT   protease poly:peptide(s)).
XX
PS   Claim 1; Col 167-182; 101pp; English.
XX
CC   This sequence is encoded by a Porphyromonas gingivalis gene of the
CC   invention. This sequence represents the hagE haemagglutinin protein. The
CC   polypeptides are used to produce antibodies to organisms associated with
CC   periodontal disease. The antibodies are also used in purification and
CC   identification procedures. The genes and polypeptides are used as
CC   vaccines against periodontal disease
XX
SQ   Sequence 1687 AA;

Query Match     86.2%;  Score 2058;  DB 2;  Length 1687;
Best Local Similarity  88.8%;  Pred. No. 5.8e-158;
Matches 389;  Conservative  12;  Mismatches  31;  Indels  6;  Gaps  4;

Qy      3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS 62
          |||||||||||||||||||||||||||||||||||||       ||   | :
Db    935 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 991

Qy     63 ASY-INFEGPQNPDNYLVTPELSLPGGGTLIFWVCAQDANYASEHYAVYASSTGNDASNF 121
          | :: |||||||   |||:| ||||||||||||||||||||||||||||||||||||||
Db    992 ESFGLGGIGVLITPDNYLITPALDLPNGGKLIFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy    122 ANALLEEVLTAKTVVTAPEAIRGTRVGGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          ||||||||||||||  ||||||||:||||||  :|:||||||||||||:|| |:|||||
Db   1052 TNALLEETITAKG-VRSPEAIRG-RIQCTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy    182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSQQLDWLTAHGGTNV 241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSQLDWLTAHGGTNV 1169

Qy    242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGTNAGDFTVV 301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGTNAGDFTVV 1229
```

PT Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
PT useful for protecting animals and humans from gingivalis and periodontal
PT diseases.
XX
PS Disclosure; Page 68-73; 95pp; English.
XX
CC The present sequence represents an arginine-specific protease of
CC Porphyromonas gingivalis. The following peptides, derived from Arg- and
CC Lys-specific high molecular weight proteases, offer protection against
CC infection: YYTYVRDGK IKEGLTATTE DDGVATGNHE YCVEKYTAGS VSPKVC (I);
CC YTPVEEKQNG RMVIVAKKY (II); QLPFIFDVAC VNGDFLFSMP CFAEALMRAQ (III);
CC GEPNPYQPVS NLTATTQGQK VTLKWDAPSTK (IV); GNHEYCVEVK YTAGVSPKVC KDVTV (V);
CC RMFMNYEPGR YTPVEEKQNG (VI); TFAGFEDTYK RMFMNYEPGR (VII); DYTYTVYRDG
CC TKIKEGLTAT TFEEDGVATG NMEYCVCVKY TAGVSPKVC (VIII); YYTYVRDGT KIKEGLTATTF
CC EEDG (IX); RDGTKIKEGL TATTFEEDGV ATGN (X); KIKEGLTATT FEEDGVATGN HEY (XI)
CC ; KWDAPNGTPN PNPN PGTTLSE (XII); and YTPVEEKENG RMIVIVAKKY
CC (XIII). They are used in vaccines to protect animals, including humans,
CC from gingivitis and/or periodontal diseases
XX
SQ Sequence 1704 AA;

Query Match   86.2%; Score 2058; DB 2; Length 1704;
Best Local Similarity 88.8%; Pred. No. 5.9e-158;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

```
Qy     3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS 62
           |||||||||||||||||||||||||||||||||||           |:|:|  ||: |
Db   952 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy    63 ASY-INFEGPQONPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          ||: :|  |||||||||||||||||||||||||||||||||||||||
Db  1009 ESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          |||||:  :|||   :  ||||| :|:|| :|||||||||||||||||
Db  1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy   182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1187 VASFSWNGMALNPDNYLIISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db  1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPDTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1307 DIQFTMGGSPTPDTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy   422 VCVNVTINPTQFNPVQNL 439
          |||||:||||:|:|
Db  1367 ECVNVTINPTQFNPVKNL 1384
```

RESULT 15
AAY67396
ID AAY67396 standard; protein; 1704 AA.
XX
AC AAY67396;
XX
DT 25-APR-2000 (first entry)
XX
DE Arg-gingipain-2 amino acid sequence.
XX
KW Arginine specific proteinase; Arg-gingipain; gingipain-2; haemagglutinin;
KW immunogenic component; vaccine; inflammatory response; tissue damage;
KW periodontal disease.
XX
OS Porphyromonas gingivalis.

Query Match   86.2%; Score 2058; DB 2; Length 1704;
Best Local Similarity 88.8%; Pred. No. 5.9e-158;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

```
Qy     3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS 62
           |||||||||||||||||||||||||||||||||||           |:|:|  ||: |
Db   952 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy    63 ASY-INFEGPQONPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          ||: :|  |||||||||||||||||||||||||||||||||||||||
Db  1009 ESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          |||||:  :|||   :  ||||| :|:|| :|||||||||||||||||
Db  1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy   182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1127 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1187 VASFSWNGMALNPDNYLIISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db  1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPDTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1307 DIQFTMGGSPTPDTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy   422 VCVNVTINPTQFNPVQNL 439
          |||||:||||:|:|
Db  1367 ECVNVTINPTQFNPVKNL 1384
```

RESULT 14
AAW34843
ID AAW34843 standard; protein; 1704 AA.
XX
AC AAW34843;
XX
DT 03-JUN-1998 (first entry)
XX
DE Arg-gingipain high molecular weight prepolyprotein sequence.
XX
KW Arg-specific gingipain protease; gingivalis; periodontal disease;
KW vaccine; infection.
XX
OS Porphyromonas gingivalis.
XX
FH Key   Location/Qualifiers
FT Protein   1..227
FT   /note= "precursor protein"
XX
PN WO9734629-A1.
XX
PD 25-SEP-1997.
XX
PF 21-MAR-1997; 97WO-US004635.
XX
PR 22-MAR-1996; 96US-0013945P.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
PA (MORE-) MOREHOUSE SCHOOL MEDICINE.
XX
PI Potempa J, Travis J, Genco C;
XX
DR WPI; 1997-479993/44.
DR N-PSDB; AAT93872.
XX

```
XX
FH  Key              Location/Qualifiers
FT  Region           229..719
FT                   /note="Amino acids 229-719 are specifically claimed"
FT  Region           720..1185
FT                   /note="Amino acids 720-1185 are specifically claimed"
XX
PN  US6017532-A.
XX
PD  25-JAN-2000.
XX
PF  08-NOV-1994;  94US-0033630B.
XX
PR  10-SEP-1993;  93US-00119361.
PR  24-JUN-1994;  94US-0026544l.
XX
PA  (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI  Potempa JS, Travis J;
XX
DR  WPI; 2000-136659/12.
DR  N-PSDB; AAZ60181.
XX
PT  New Porphyromonas gingivalis arginine-specific protease preparation
PT  useful for preparing vaccines against periodontal disease and for
PT  screening for Arg-gingipain inhibitors.
XX
PS  Claim 1; Col 29-42; 55pp; English.
XX
CC  This sequence represents a Porphyromonas gingivalis arginine-specific
CC  proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
CC  Gingipain-2 consists or a 50kD protease component non-covalently
CC  associated with a 44kD haemagglutinin component. The proteinase is
CC  stimulated by glycine containing peptides and glycine analogues. It is
CC  inhibited by cysteine protease group specific inhibitors. The protease
CC  preparation can be used in immunogenic compositions and vaccines against
CC  inflammatory response and tissue damage caused by P. gingivalis in
CC  periodontal disease. It can also be used to screen for agents that
CC  modulate Arg-gingipain proteinase activity inhibitors
XX
SQ  Sequence 1704 AA;
```

```
Query Match      86.2%;  Score 2058;  DB 3;  Length 1704;
Best Local Similarity   88.8%;  Pred. No. 5.9e-158;
Matches  389;  Conservative  12;  Mismatches  31;  Indels  6;  Gaps  4;

Qy     3  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICVSS  62
          ||||||||||||||||||||||||||||||||||||||||  ||:||  ||:|
Db   952  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS  1008

Qy    63  ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF  121
          ||:|:                ::|:||||||||||||||||||||||||||||||||||
Db  1009  ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF  1068

Qy   122  ANALLEBVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD  181
          |||::|:|:   :  ||||||||| ||:||::| |||||:::|:||||:|  :|: ||:|
Db  1069  TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD  1126

Qy   182  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV  241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1127  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV  1186

Qy   242  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1187  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  1246

Qy   302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1247  FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  1306

Qy   362  DIQFTMGGSPTPTDYTTVYRDGTKIIKEGLTETTFEBDGVATGNHEYCVEVKYTAGVSPK  421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1307  DIQFTMGGSPTPTDYTTVYRDSTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  1366
```

```
Qy   422  VCVNVTINPTQFNPVQNL  439
          ||||||||||||||||:||
Db  1367  ECVNVTINPTQFNPVKNL  1384
```

```
Search completed: May 18, 2004, 11:42:44
Job time : 46.2755 secs
```

```
GenCore version 5.1.6
Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May 18, 2004, 11:37:00 ; Search time 11.4621 Seconds
                                       (without alignments)
                                     3684.135 Million cell updates/sec

Title:           US-08-570-311-22
Perfect score:   2388
Sequence:        1 GTPNPNPNPGTTLSESF........PKVCVNVTINPTQFNPVQNL 439

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:     283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR 78:*
                 1: Pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	Query Match*	Length	DB	ID	Description
1	2388	100.0	2628	2	T28651	hemagglutinin A -
2	2058	86.2	2704	2	A55426	gingipain R (EC 3.
3	2037	85.3	1526	2	S49763	gingipain R (EC 3.
4	2025.5	84.8	1732	2	T30836	lysine-specific cy
5	251.5	10.5	991	2	I40229	arginyl endopeptid
6	136.5	5.7	1052	2	AF2959	conserved hypothet
7	136.5	5.7	1341	2	H98323	hypothetical prote
8	129.5	5.4	1248	2	C89874	autolysin [importe
9	129	5.4	691	2	B75522	hypothetical prote
10	124	5.2	5188	2	B85547	probable RTX famil
11	124	5.2	5291	2	F90696	hypothetical prote
12	123.5	5.2	1684	2	S10789	amylase A-180 - al
13	123	5.2	596	2	A55976	cellulose 1,4-beta
14	122.5	5.1	1090	2	S59077	cellulose 1,4-beta
15	121	5.1	1649	2	C86822	hypothetical prote
16	120.5	5.0	1345	2	H90975	hypothetical prote
17	119.5	5.0	2817	2	B97033	uncharacterized pr
18	119	5.0	715	2	JC4908	alkaline serine pr
19	118.5	5.0	1904	2	T13256	tail-host specific
20	118	4.9	1441	2	A86685	prophage pil specific
21	118	4.9	1939	2	D97316	probable S-layer p
22	118	4.9	4199	2	S76412	hypothetical prote
23	117.5	4.9	607	2	C69503	conserved hypothet
24	117	4.9	1122	2	T18346	MGC1 protein precu
25	116.5	4.9	1274	2	T10729	transferrin-like p
26	116.5	4.9	2660	2	E85822	probable invasin Z
27	116	4.9	4936	2	AH2515	hypothetical prote
28	115.5	4.8	702	2	S48753	major surface prot
29	114.5	4.8	2468	2	A83412	hypothetical prote
30	114	114	4.8	1461	2 E90696	hypothetical prote
31	114	114	4.8	1461	2 A85547	hypothetical prote
32	112	112	4.7	872	2 S49541	cellulase - Cellul
33	112	111.5	4.7	1377	2 I54632	tsh protein - Esch
34	111.5	111.5	4.7	702	2 S48754	major surface prot
35	111.5	111.5	4.7	882	2 H82754	family 3 glycoside
36	111	111	4.6	465	2 A47023	S-layer protein -
37	111	111	4.6	3624	2 AD0835	large repetitive p
38	110.5	110.5	4.6	713	2 B75489	hypothetical prote
39	110.5	110.5	4.6	1268	2 AB0204	conserved hypothet
40	110.5	110.5	4.6	2044	2 AB1180	probable peptidogl
41	110.5	110.5	4.6	2554	2 AB3528	extracellular seri
42	110	110	4.6	987	2 A64474	hypothetical prote
43	109.5	109.5	4.6	891	2 T36423	probable large, mu
44	109.5	109.5	4.6	938	2 AF1772	internalin-like pr
45	109.5	109.5	4.6	1034	2 T30551	beta-galactosidase

ALIGNMENTS

```
RESULT 1
T28651
hemagglutinin A - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C;Accession: T28651
R;Han, N.; Whitlock, J.; Progulske-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A;Title: The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381 contains four
A;Reference number: Z20494; MUID:97047672; PMID:8926061
A;Accession: T28651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2628 <HAN>
A;Cross-references: EMBL:U41807; NID:g1552410; PID:g1469916; PIDN:AAB17128.1
C;Genetics:
A;Gene: hagA

Query Match        100.0%;   Score 2388;   DB 2;   Length 2628;
Best Local Similarity   100.0%;   Pred. No. 6.2e-159;
Matches 439;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDNNWTTTPPGGTSFAGHNSAICV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1862 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDNNWTTTPPGGTSFAGHNSAICV 1921

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1922 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVVASSTGNDASN 1981

Qy  121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1982 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 2041

Qy  181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2042 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 2101

Qy  241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2102 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 2161

Qy  301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2162 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 2221

Qy  361 DDIQFTMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2222 DDIQFTMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 2281

Qy  421 KVCVNVTINPTQFNPVQNL 439
          |||||||||||||||||||
```

Db   2282 KVCVNVTINPTQFNPVQNL 2300

RESULT 2
A55426
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
N;Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP; R
C;Species: Porphyromonas gingivalis
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
C;Accession: A55426; D53113
R;Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr, F
J. Biol. Chem. 270, 1007-1010, 1995
A;Title: Molecular cloning and structural characterization of the Arg-gingipain proteina
A;Reference number: A55426; MUID:95138080; PMID:7836351
A;Accession: A55426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1704 <PAV>
A;Cross-references: GB:U15282; NID:g557067; PIDN:AAA69539.1; PID:g557068
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: D53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 228-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141694)
C;Keywords: cysteine proteinase; hydrolase

Query Match   86.2%;   Score 2058; DB 2; Length 1704;
Best Local Similarity   88.8%;   Pred. No. 4.9e-136;
Matches 389;   Conservative 12;   Mismatches 31;   Indels   6;   Gaps   4;

Qy   3 PNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS 62
           |||||||||||||||||||||||||||||||||||| ::|| ||||||
Db   952 PNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          ||| |||||||||||||||||||| | |||||||||||||||||||||||||||||||||
Db 1009 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVGGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          |||||| :||| ||||||||| : ||||||||||||||||||||||||| :: :||
Db 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy   182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1127 EVEIKANGKRADFTETFESSTHGEATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1307 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy   422 VCVNVTINPTQFNPVQNL 439
          ||:||||||||||||:||
Db 1367 ECVNVTINPTQFNPVKNL 1384

RESULT 3
S49763
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis (fragment)
C;Species: Porphyromonas gingivalis
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997

C;Accession: S49763
R;Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.
submitted to the EMBL Data Library, November 1994
A;Description: Cloning, sequence analysis and expression in Escherichia coli of prpR1 o1
A;Reference number: S49763
A;Accession: S49763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <ADU>
A;Cross-references: EMBL:X82680
C;Genetics:
A;Gene: prpR1
C;Keywords: cysteine proteinase; hydrolase

Query Match   85.3%;   Score 2037; DB 2; Length 1526;
Best Local Similarity   87.9%;   Pred. No. 1.2e-134;
Matches 385;   Conservative 13;   Mismatches 34;   Indels 6;   Gaps   4;

Qy   3 PNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS 62
           |||||||||||||||||||||||||||||||||| ::|| ||||||
Db   954 PNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1011 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVGGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          |||||| :||| |||||||| : ||||||||||||||||||||||||| :: :||
Db 1071 TNALLEETITAKG-VRSPEAIRG-RIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy   182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1129 EVEIKANGKRADFTETFESSTHGEATAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1188

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1249 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy   362 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1309 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy   422 VCVNVTINPTQFNPVQNL 439
          ||:|||||||||:|
Db 1369 KCVNVTVNSTQFNPVKNL 1386

RESULT 4
T30836
lysine-specific cysteine proteinase porphypain (EC 3.4.22.-) - Porphyromonas gingivalis
N;Alternate names: lysine-specific cysteine proteinase 1, 60K
C;Species: Porphyromonas gingivalis
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30836; T30837; T30526; A53113
R;Barkocy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulske-Fox, A.; Lantz
J. Bacteriol. 178, 2734-2741, 1996
A;Title: Analysis of the prtP gene encoding porphypain, a cysteine proteinase of Porphy1
A;Reference number: Z20895; MUID:96213011; PMID:8631659
A;Accession: T30836
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1732 <BAR>
A;Cross-references: EMBL:U42210; NID:g1314325; PID:g1314326; PIDN:AAB06565.1
R;Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z20896
A;Accession: T30837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA

A;Residues: 1-795,'I',797-1389,'N',1391-1478,'Y',1480-1732 <SLA>
A;Cross-references: EMBL:U75366; NID:g2182811; PID:g2182812; PIDN:AAB60809.1
R;Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A;Title: IS195, an insertion sequence-like element associated with protease genes in Por
A;Reference number: Z20844; MUID:9829016; PMID:9632563
A;Accession: T30526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA

A;Residues: 1-795,'I',797-1389,'N',1391-1478,'Y',1480-1732 <SLA>
A;Cross-references: GB:D26470; NID:g927644; PIDN:BAA05484.1; PID:g927645

Query Match   10.5%;   Score 251.5; DB 2;   Length 991;
Best Local Similarity   25.9%;   Pred. No. 1.1e-09;
Matches 120;   Conservative   50;   Mismatches 171;   Indels 123;   Gaps 21;

Qy   8 NPNPGTTLSESFENGIPASWKTIDADGDG--NNWT-----------------------TPPP 46
             ||| | |||||| ||| |
Db   529 HPNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQ 588

Qy   47 GGTSFAG------HNSAICVSSASYINFEGPQNPDNYIVTPELS-LPGGGTLTFWVCAQD 99
                          :: : :|| :| | || | |
Db   589 INLTDASVNVSCDYNGAIATISANGKMF-GSAVVENGTATINLTGLTNESTLTITV---- 643

Qy   100 ANYASEHYAVYASSTG--NDASNFANALLEEVLTAKIVVTAPEAIRGTRVQGTWYQKTVQ 157
             ::   :| :: | | | :   :: :   :
Db   644 VGYNKETVIKTINTNGENPYQPVSN------LTATT--------QGQKVTLKWDAPSTK 689

Qy   158 LPAGTKYVAFRHFGCTDFFWINL-DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDAD 216
            |||| | | |||| :||| | : || | | |
Db   690 TNATTN-TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWNDGS--GYQILLDAD 746

Qy   217 GDGQGWLCLSSGQLDWLTAHGGTNVVASFSWN------GMALNPDNYLISKDVTGATKVKY 271
            | | | : ||| | :|| :| : :: :
Db   747 HDQYGVQVIPSDTHTLWPNCSVPANLFAPFEYTVPENADPSCSPTNMIM--DGTASVNI-- 802

Qy   272 YYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEADGAKPQS-- 329
             :   : |
Db   803 ------------PAGTY--------------------------DFAI------------AAPQANA 818

Qy   330 -VWI------ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTYVY 381
             :| ||| ||||| : || | ||| |||
Db   819 KIWIAGGQPTKEDDYVFEAGKKY----HFLMKKMGSGDGTELTISEGGG--SDYTYTYV 871

Qy   382 RDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVN 425
            |||||||||||||||| :: : : ||| |||| ||||| |
Db   872 RDGTKIKEGLTETTYRDAGMSAQSHEYCVEVKYAAGVSPKVCVD 915

RESULT 6
AF2959
conserved hypothetical protein Atu3276 [imported] - Agrobacterium tumefaciens (strain C
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 18-Nov-2002
C;Accession: AF2959
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF2959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1052 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44092.1; PID:g17741659; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
C;Gene: Atu3276
A;Map position: linear chromosome

Query Match   5.7%;   Score 136.0; DB 2;   Length 1052;
Best Local Similarity   21.1%;   Pred. No. 0.14;
Matches 119;   Conservative   61;   Mismatches 148;   Indels 235;   Gaps 33;

Qy   13 TTTLSESFENGIPASWKTIDADGDGNWTT-----------TPPPGGTSFAGHNSAICVSSAS 64
             ||||||||||||||||||||| ||| ||| || | || |
Db   52 TVTSGEAIIGRHSASEATV--TGDGSKWTTGDLQVGGDTSDPGG--LAGNGT--------- 99

RESULT 5
I40229
arginyl endopeptidase - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
C;Accession: I40229
R;Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A;Title: Structural characterization of arginipain, a novel arginine-specific cysteine
A;Reference number: I40229; MUID:95168884; PMID:7864651
A;Accession: I40229

A;Residues: 1-1350,'N',1352-1363,'Y',1365-1447,'H',1449-1732 <LEW>
A;Cross-references: EMBL:AF017059; NID:g2738802; PID:g2738803; PIDN:AAC26523.1
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isola
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: A53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 229-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141690)
C;Genetics:
A;Gene: prtP; prtK
C;Keywords: cysteine proteinase; hydrolase

Query Match   84.8%;   Score 2025.5; DB 2;   Length 1732;
Best Local Similarity   87.1%;   Pred. No. 9.4e-134;
Matches 384;   Conservative   16;   Mismatches 34;   Indels 7;   Gaps 5;

Qy   1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
            ||||||||||| ||||||||||||||||||||||||| : ||||| |||||||||
Db   969 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy   60 VSSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
            |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1026 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy   119 SNFANALLEEVLTAKIVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
            |||| ||||| ||| ::||||||||| :::||:||:||||||||||||||||||||||
Db 1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy   179 NLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 238
            :|||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db 1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGDGGWLCLSSGQLDWLTAHGG 1203

Qy   239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy   299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
            |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db 1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy   359 LLDDIQFTMGGSPTPTDYTYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1324 LLDDIQFTMGGSPTPTDYTYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy   419 SPKVCVNVTINPTQFNPVQNL 439
            |||||| ||:||||||||||||
Db 1384 SPKKCVDVTVNSTQFNPVQNL 1404

```
Qy    65  YINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGN-------  116
              :||  |||| ||||||:|:|||||||||||||||||||||||||||||||||
Db   100  ------------------DSTVA--HLGVVAGATGSAIVDGKGS             132

Qy   117  ----DASNF------ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVA  166
              ||:||          :|:||||||||||||||||:|:::||||||||||||||||
Db   133  VWTVDRNSLEVGVSGAGSL--AVTGGGLVDAANIIGINTGG--NGSVRVSGADSTVK   186

Qy   167  FRHFGCTDFFWINL-----DEVEIKANG----KRADFTETFESSTHGEAPAEWTTIDADGD 218
              ||||||||||||      ||:||||     ||||||||||||||||||||||||||||
Db   187  SR-----SD--LNVGLYGNGSMTVERGGAVKSRDGVVATYGGST-----SAVTVTGD     231

Qy   219  GQGWLCLSSGQLDWLTAHGGT-NVVASFSWNGMALNPDNYLI-----SKDVTGA-TK    268
              |||||||||||||||||||| ||||||||||||||||||||     ||||||| ||
Db   232  GSSWAMTGTFFVGY--ASGATGNVTVS---NGGAIRATGVTLGDLAGASGTMTITGAGSK  286

Qy   269  VKYYY--------DHY-------------AVMISKTGT---             293
              |||||        |||             |||||||||
Db   287  VTAYVDNGTVNSGSVDVGFQGSGSLSVVNGGSLDAYNLYVGNALGSSGAVLVSGVGSHVS  346

Qy   294  ------NAGD------------FTVVFEETPNGINKGGARFGLSTEADGAKPQSV - 330
              :|||            |||||||||||||||||||||||||||||||||||| :
Db   347  VDGLMVVGNAGNGSVEITGGASLAAPTILIATEAGSTGVLSIGAGSGQTARSAGA - 401

Qy   331  WIE-RTVDLPAGTKYVAFRHYNCSDLNYILLDDIQ-----             378
              ||| |||||||||||||||||||||||||||||||
Db   402  -VEARAIAFGAGNGSIVFNH---SETGYTLSADISGAGRVVAEAGVTTLSGNNS------- 451

Qy   379  TVYRDGTKIKEGLTETT----                   402
              ||||||||||||||||
Db   452  --YSGGTTISAGMLKGTAKSFGSGGIVNNAELVVDGGGTLSNAISGTGSFEKTGDGNLLL  509

Qy   403  TGNHEYCVEVKYTAGVSPKVCVN 425
              |||||||||||||||||||||||
Db   510  TGNSTYSGATAVSAG---KLSVN 529
```

RESULT 7

H98323

hypothetical protein AGR_L_3085 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: H98323
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
C.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1341 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK90114.1; PID:g15160106; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_3085
A;Map position: linear chromosome

```
Query Match             5.7%;     Score 136.5;  DB 2;     Length 1341;
Best Local Similarity  21.1%;     Pred. No. 0.19;
Matches  119;  Conservative    61;  Mismatches  148;  Indels 235;  Gaps 33;

Qy    13  TTTLSESFENGIPASWKTIDADGDGNNWTT--------TPPPGGTSFAGHNSAICVSSAS   64
              :||||||||||||||||||||||||          ||  ||||||||
Db   341  TVTSGEAIIGRHSASEATV--TGDGSKWTTGDLQVGGDTSDPGG--LAGNGT--------  388

Qy    65  YINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGN-------  116
              :||  |||| ||||||:|:|||||||||||||||||||||||||||||||||
Db   389  ------------------DSTVA--HLGVVAGATGSAIVDGKGS             421

Qy   117  ----DASNF------ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVA  166
              ||:||          :|:||||||||||||||||:|:::||||||||||||||||
Db   422  VWTVDRNSLEVGVSGAGSL--AVTGGGLVDAANIIGINTGG--NGSVRVSGADSTVK     475
```

RESULT 8

C89874

autolysin [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: C89874
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: C89874
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1248 <KUR>
A;Cross-references: GB:BA000018; PID:g137000854; PIDN:BAB42150.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: atl

```
Query Match             5.4%;     Score 129.5;  DB 2;     Length 1248;
Best Local Similarity  21.9%;     Pred. No. 0.53;
Matches  103;  Conservative    47;  Mismatches  167;  Indels 153;  Gaps 25;

Qy     3  PNPNPNPNPGTT---LSESFENGIPASWKTIDADGDG--NNWTTTPPP-             46
              :|||||||||    ||||||||||||||||||||||  :||||||
Db   586  PTPTPIPKPSTPTTNNKLITVSSLNGV----AQINAKQNGLFTTVVDKTGKPTKEVQKTFA 641

Qy    47  -------GGTSF--AGHNSAICV-----SSASYINFEGPQN-PDNYLVTPELSL------  85
              :|||| |||||||      ||||||||||| ||||||||||||||
Db   642  VTKEASLGGNKFYLVKDYNSPTLIGWKQGDVIYNNAKSPVNVMQTYTVKPGTKLYSVPW  701

Qy    86  ------PGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFAN-------ALLE     127
              ||||||||||||||||||||||||||||||:|||     ||||
Db   702  GTYKQEAGAVSGTGNQTFKATKQQQIDKS-----IYLFGTVNGKSGWVSKAYLAVPAAPKK 757

Qy   128  EVLTAKT-----VVTAPEAIR----GTRVQGTWYQKTVQLPAGTKYVAFRHFG    171
              ||||||      :||||| |     |||||||||||||||||||||
Db   758  AVAQPKTAVKAYTVKPQTQTVSKIAQVKPNNTGIRASVVEKTAK--NGAKY-ADRTFY  814

Qy   172  CT-----             DFFWINLDEVEIKANGKRADFTE-TFESSTHGEAPAE    209
              ||                  |||||:||:|||||  :|: ||  ||| ||
Db
```

```
Db  815  VTKERAHGNETYVLLNNTSHNIPLGWFNVKDLNVQNLGKEVKTTQKYTVNKSNNGLSMVP  874
Qy  210  WTTIDADGDGQWLCLSSGQLDWLTAHGGTNVVASFSWN------GMALNPDNYLISKDV   263
Db  875  WGT-------KNQVILTGNNI----AQGTFNATKQVSVGKDVVLYGTINNRTGWVNAKDL  923
Qy  264  TGATKVK---------YYYAVNDGFPGDHYAVMISKTGT------NAGDFTVFEETPN    307
Db  924  TAPTAVKPTTSAAKDYNTYVIKNG-NGYYYVTPNSDTAKYSLKAFNEQFPAVVKEQVIN  982
Qy  308  G------INKGGARFGLSTEADGAK------PQSWIERTVDLPAGTKY            344
Db  983  GQTWYYGKLSNGKLAWIKST--DLAKELLKYNQTGMTLNQVAQIQAGLQY         1030
```

RESULT 9
B75622
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: B75622
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75622
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-691 <WHI>
A;Cross-references: GB:AE001826; NID:g6460827; PIDN:AAF12628.1; PID:g6460924; TIGR:DRB0C
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRB0037
A;Map position: megaplasmid
A;Genome: plasmid
A;Note: plasmid MP1

Query Match   5.4%; Score 129; DB 2; Length 691;
Best Local Similarity 21.3%; Pred. No. 0.27;
Matches 113; Conservative 52; Mismatches 192; Indels 174; Gaps 24;

```
Qy    4  NPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTP-----------PPGGTSF   51
Db  230  SPNPLPPTGTTT----------PGTGTPVGSGNGPAESTGPAYTGIGSTGSGTQPPVVVDA 280
Qy   52  AGHNSAICVSSASYINFEGP------QNPD---NYLVTPELS-LPGGGTLTFWVCAQDA  100
Db  281  SGNQYAYPKADAETVNPDSVTMTTIITNPNPTAATYELVPDLSGLPAGVTVTF---TDA   336
Qy  101  N-----YASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQ   153
Db  337  NGNPLPDTDGNGRPEVTAGPNGGTA-------TYRVVVTYPDTESAAAVAG----      380
Qy  154  KTVQLPAGTK---------YVAFRHFGCTDFFWINLDEVEIKANGKRADFT--       195
Db  381  -PIRIPVGVDGNRDGIVDATVTYNVLLSNLKFGNTNGTALGVSDVPVTRTVTQPSVTTAV  439
Qy  196  ---ETFESSTH----------GEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGT    239
Db  440  VFPMDLFNDGAYDGNYALSGSTPIGPVKYYATNPDTDGDG--VLSPAELAALPAE---   492
Qy  240  NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGTNAGDFT  299
Db  493  --IAS---------TGAVPVKTEKTV-----YAVVTIPAGQAPGDYM             523
Qy  300  VVFEETPNGINKGGARFGLSTEADGAKPQ-SVWIERTVDLPAGT                342
Db  524  VT--QTATGSLSGTTKSFNTDKVTVTSPNGSLLIAKRVTTPGTTTPSLNATANPGDAVSY  581
Qy  343  KYVAFRHYNCSDLNYILLDDIQFTMG--------GSPTPTDYTYTV---------YRD-  383
Db  582  TVTATNYNTSLYGLVLRDPSSNNLGSFSSNNVFGFIKPSSLRATVSGVSGATVLYRTSN  641
Qy  384  -GTKIKEGLTE--TTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTINPT         431
Db  642  LNTWAAQPTVDANTWVEVGVDTNNNS--QIDSGDVFPNAVITLTLQGT            689
```

RESULT 10
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B85547
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85547
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5188 <STO>
A;Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:ZC
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0615

Query Match   5.2%; Score 124; DB 2; Length 5188;
Best Local Similarity 23.9%; Pred. No. 8.4;
Matches 138; Conservative 63; Mismatches 189; Indels 188; Gaps 35;

```
Qy    10  NPGTTTL-SESFENGIPAS------------WKTIDA--DGDGNNWTTT--          PPPG    47
Db  2180  NTFFTVLDASGNWSVGVPASVVSALANGTVTINASVTDAGGNSGSATHQVTVNTGLPTIT 2239
Qy    48  GTSFAGHN-----------SAICVSSASYINFEGPQ-----NPDNYLVTPELSLPGGGTLTFW  94
Db  2240  FNAISGDNIILNADEKGQPLTISGGSTGLATGAQVTVTLNGHNYSATTDAS--GNWLTLTVP 2297
Qy    95  VCAQDANYASEHYAVYASST--GNDASNFANALLEEVLTAKTVVT------A          138
Db  2298  V-SDIAALGQANYTVSASATSAAGNTASSQANLLVDSGLPDVTINTVAGDDIINAAEAGA  2356
Qy   139  PEAIRG--TR--------VQGTWYQKTVQLPAGTKYVAFRHFGCTDFFW-INLDEVEI     185
Db  2357  DQTISGVVTRAAAGDTVTVLGGNYTATVQ-----------SNLSWSVSVPTADL       2401
Qy   186  KANGKRADFTETFESSTHGEAPAEWT---TIDAD--------GD------GQ----      220
Db  2402  QALG-NGDLTITASVTNANGNTGSGTRDITTIDANLPGLRVDTVAGDDIVNSIEHGQALVI 2460
Qy   221  --GWLCLSSGQLDWLTAHG--GTNVVASFSWN-GM-ALNPDNY---LISKDVTGATKVK   270
Db  2461  TGGSSGLNAGAVLTVTINSVAYSATVQADGSWSVGIPAANVSAWPAGPLTVEVDQQSSAN  2520
Qy   271  YYYAVNDGFPGDHYAVMIS-------------KTGTNAGDFTVVFEETPNGINK----   311
Db  2521  NPVSVSHPFTVDLTAVAISINTVASDDVINAAEKGTN------LTLSGSTSGIESGQTVTV 2575
Qy   312  --GGARFGLSTEADGAKPQSWIERTVDLPAGTKYV------------AFRHYNCS--    353
Db  2576  TFGGKTYTASVAANGSWSVNVPAADLATLPEGAANVQASVSSASGNSASATHAYSVDASA  2635
Qy   354  ---DLNYILLDDI-------QFTMGGSPT-PTDYTYTVVRDGTKIKEGLTET-TFEEDG   400
Db  2636  PTLTINTIASDDILNAAEAEGSPLTISGTSTAETGQTVTVTLNGA------TYTGVQADG  2689
Qy   401  VATGNHEYCVEVKYTA-GV--SPKVCVNVTINPTQFNP                      435
Db  2690  ------SWSVSVPTSALGALNASNTVSAATVNDKAGNP                     2721
```

RESULT 11

F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: F90696
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90696
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5291 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33965.1; PID:g13360000; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0542

Query Match   5.2%;   Score 124;   DB 2;   Length 5291;
Best Local Similarity   23.9%;   Pred. No. 8.6;
Matches 138; Conservative 63; Mismatches 189; Indels 188; Gaps 35;

Qy   10 NPGTTTL--SESFENGIPAS----------WKTIDA--DGDGNNWTTT------------PPPG 47
                ||||| : || ::| |||: || ||||:||||| ||||
Db   2180 NTFTTVLDASGNWSVGVPASVVSALANGTVTINASVTDAGGNSGSATHQVTVNTGLPTIT 2239

Qy   48 GTSFAGHN----------SAICVSSASYINFEGPQ-----NPDNYLVTPELSLPGGGTLTFW 94
               || ||| |||:|||| || ||||| ||||:| :|| :||||| || |
Db   2240 FNAISGDNILNADEKGQPLTISGGSTGLATGAQVTVTLNGHNYSATTDAS--GNWTLTVP 2297

Qy   95 VCAQDANYASEHYAVYASST--GNDASNFANALLEEVLTAKTVVT-------------A 138
                ||| |||| ||:| |||: |||:||||||||||| | : | |
Db   2298 V-SDLAALGQANYTVSASATSAAGNTASSQANLLVDSGLPDVTINTVAGDDIINAAEAGA 2356

Qy   139 PEAIRG--TR----------VQGTWYQKTVQLPAGTKYVAFRHFGCTDFFW--INLDEVEI 185
                 ||:| || | |||| :||| ::| :|:|||:
Db   2357 DQTISGVVTRAAAGDTVTVTLGGNTYTATVQ------------SNLSWSVSVPTADL 2401

Qy   186 KANGKRADFTETFESSTHGEAPAEWT---TIDAD----GD-----GQ------ 220
               |||||||||||| ||:| :|| :| | ||| || ||
Db   2402 QALG-NGDLTTTASVTNANGNTGSGTRDITIDANLPGLRVDTVAGDDIVNSIEHGQALVI 2460

Qy   221 --GWLCLSSGQLDWLTAHG---GTNVVASFSWN-GM-ALNPDNY--LISKDVTGATKVK 270
                  |::|||| | |||| || |||||||| | |||||| :||||||||||:
Db   2461 TGGSSGLNAGAVLTVTINSVAYSATVQADGSWSVGIPAANVSAWPAGPLTVEVDGQSSAN 2520

Qy   271 YYYAVNDGFPGDHYAVMIS---------------KTGTNAGDFTVVFEETPNGINK------ 311
               :|| |:||| | |||: | |||:||||| ::||||| |||
Db   2521 NPVSVSHPFTVDLTAVAISINTVASDDVINAABKGTN-----LTLSGSTSGIESGQTVTV 2575

Qy   312 --GGARFGLSTEADGAKPQSVWIERTVDLPAGTKYV----------AFRHYNCS--- 353
                 |||||| |||||||||| : :||:||||||| | || ||||
Db   2576 TFGGKNTYTASVAANGSWSVNVPAADLATLPEGAANVQASVSSASGNSASATHAYSVDASA 2635

Qy   354 ---DLNYILLDDI-------QFTMGGSPT-PTDYTYTVYRDGTKIKEGLITET-TFEEDG 400
                  |:| ::|| |::|||| | |||||||||||||| :|||: :| |||||
Db   2636 PTLTINTIASDDILNAAEAGSPLTISGTSTAETGQTVTVTLNGA---------TYTGTVQADG 2689

Qy   401 VATGNHEYCVEVKYTA-GV--SPKVCVNVTINPTQFNP 435
              | ||||:| |||||| | |||||||||:|| |||
Db   2690 Q-----SWSVSVPTSALGALNASNYTVSATVNDKAGNP 2721

RESULT 12
S10789
amylase A-180 - alkaliphilic eubacterium 163-26
C;Species: alkaliphilic eubacterium 163-26
C;Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 15-Oct-1999
C;Accession: S10789
R;Candussio, A.; Schmid, G.; Boeck, A.
Eur. J. Biochem. 191, 177-185, 1990
A;Title: Biochemical and genetic analysis of a maltopentaose-producing amylase from an a

A;Reference number: S10789; MUID:90336627; PMID:1696201
A;Accession: S10789
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1684 <CAN>
A;Cross-references: EMBL:X53373; NID:g48305; PIDN:CAA37453.1; PID:g48306

Query Match   5.2%;   Score 123.5;   DB 2;   Length 1684;
Best Local Similarity   20.7%;   Pred. No. 2.1;
Matches 100; Conservative 52; Mismatches 183; Indels 149; Gaps 27;

Qy   23 GIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV-SSASYINFEGPQNPDNYLVTP 81
              ||: ||||||||| ||:|||||: :| :|: ||||||| ||||:||
Db   204 GLPRDWTPNQA--QGQNWHT--------HNDIMNKDNEAAWANWGSD--------- 241

Qy   82 ELSLPGGGTLTFWVCAQD-ANYASEHYAVYASSTGNDASNFANALLEEVLTAKTV-VTAP 139
               |:|||||||| || | |||||||| |||:|||||:||||||||::|||:|| |||
Db   242 --------------WIRADETAGY--------DNCGGSEQTMCIGFLPDIKTEVTTGVDLP 280

Qy   140 EAIRGT-RVQGTWYQKTVQLPAGTKYVAFRHFGCTDFF--WINLDEVEIKANGKRADFTE 196
               ||:|| :| : |:| : | | : :: :   :||||||:|||| |||||:|
Db   281 PILRNKWNDQASGYEDWF-VPAAEPYRQDLNIAPKDYLIKWIITSWVEBFGIDGFRVDTAK 339

Qy   197 TFESSTHGEAPAE----------WTTIDADGDGQGWLCLSSGQLDWLTA----HG---- 237
               ||:|| || | |||| | ||| |||:|||||| ||| ||
Db   340 HVEIERWAELKNEAEVALQTWRENNPDKPGANW-----DDNFWMTAEVFGHGLGKSEYFD 394

Qy   238 -GTNVVASFSW--------------------NGMALNPDNYLISKDVTGATKVKYYYAVNDGFP 280
               |||||||| |||:|||||| |||||| ||| |||| |||
Db   395 FGPDSVINFEFQNANFNNLEGLFSRYANSINTDPDFNMLSVVSSHDTKL--YSRDD----- 448

Qy   281 GDHYAVMISKTGTN-----------AGDFTVVF-EETPNGINKGGARFGLSTEADGAKPQSVW--IE 333
               |||:|: :||:|| || :| |: ||| ||| :|
Db   449 ---------LIQAGTALLLLPGGVQVFYGDETARPLGDGG--------SDPEQGTRSSMNWANIN 496

Qy   334 RIVDLPAGTKYVAFR--------HYNCSDLNYI-------LLDDIQFTMGGSPTPT 374
               |:|||||||| || | ::|:| :::|| :|||||| |
Db   497 QNV-LSHWQKLGQFRNNHIAIGAGAHQKLSDSPYTFARTYESDDIVDEVVVATGAQGTA 555

Qy   375 DYTYTVVRDGTKIKEGLT--ETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTINPTQ 432
               ||||||||||| ||||:   ::|||||||||||||| |||||||| ||| |||||||||
Db   556 VTVEGVFEDGTVVRDAYTGDETTVTK-GTAT-------FTAGTQGIILIENTAEPVT 604

Qy   433 FNPV 436
               ||||
Db   605 NLPI 608

RESULT 13
A55976
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Thermomonospora fusca
C;Species: Thermomonospora fusca
C;Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 22-Oct-1999
C;Accession: A55976
R;Zhang, S.; Lao, G.; Wilson, D.B.
Biochemistry 34, 3386-3395, 1995
A;Title: Characterization of a Thermomonospora fusca exocellulase.
A;Reference number: A55976; MUID:95186496; PMID:7880834
A;Accession: A55976
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-596 <ZHA>
A;Cross-references: GB:U18978; NID:g664822; PIDN:AAA62211.1; PID:g664823
C;Superfamily: bacterial cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;40-140/Domain: bacterial cellulose-binding domain homology <BCB>
F;41-139/Disulfide bonds: #status predicted

Query Match   5.2%;   Score 123;   DB 2;   Length 596;
Best Local Similarity   23.2%;   Pred. No. 0.58;
Matches 69; Conservative 35; Mismatches 112; Indels 82; Gaps 14;

Qy   3 PNPNPNPNPGTTLSESFENGIPASWKT------IDADGDGNNWTTTP--------PPGGT 49

RESULT 14

S59077

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) precursor - Cellulomonas fimi

N;Alternate names: exo-beta-1,4-cellobiohydrolase
C;Species: Cellulomonas fimi
C;Date: 15-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 22-Oct-1999
C;Accession: S59077; S59044; I40694
R;Shen, H.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.J.
Biochem. J. 311, 67-74, 1995
A;Title: Cellobiohydrolase B, a second exo-cellobiohydrolase from the cellulolytic bacte
A;Reference number: S59044; MUID:96003898; PMID:7575482
A;Accession: S59077
A;Molecule type: DNA
A;Residues: 1-1090 <SHE>
A;Cross-references: EMBL:L38827; NID:g790696; PIDN:AAB00822.1; PID:g790698
A;Accession: S59044
A;Molecule type: protein
A;Residues: 54-78;456-461 <SHW>
R;Shen, H.; Tomme, P.; Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Warren, R.A.; Miller, R.
Biochem. Biophys. Res. Commun. 199, 1223-1228, 1994
A;Title: Stereochemical course of hydrolysis catalysed by Cellulomonas fimi CenE, a memb
A;Reference number: I40694; MUID:94197708; PMID:8147863
A;Accession: I40694
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 54-106 <RES>
A;Cross-references: GB:L29042; NID:g556035; PIDN:AAA50257.1; PID:g556036
C;Genetics:
A;Gene: cbhB; cenE
C;Superfamily: bacterial cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-53/Domain: signal sequence #status predicted <SIG>
F;54-1090/Product: cellulose 1,4-beta-cellobiosidase #status experimental <MAT>
F;989-1090/Domain: bacterial cellulose-binding domain homology <BCB>
F;990-1089/Disulfide bonds: #status predicted

Query Match   5.1%;   Score 122.5;   DB 2;   Length 1090;
Best Local Similarity   21.9%;
Matches 102;   Conservative 47;   Mismatches 157;   Indels 159;   Gaps 25;

```
Qy    26 ASWKT---IDADGDGNNWTTPPPGGTSFAGHNSAICVSSASYINFEGPQNPDNLVTPE 82
             :: ||||||||||||          :::: :::::: ::: :: ::
Db   515 ASWKVPSELKWTGKPDTWNAAAPTG-----NPGLTVEVTSY----------------- 550

Qy    83 LSLPGGGTLTFWVCAQDANYASE--HYAVYASSTGNDAS-NFANALLEEV-------LT 131
             |||| :: ||  :::::       :::  :::::: ::    ::::
Db   551 -------GQDVGVAADTARALLFYAAKSGDTASRDKAKALLDAIWANNQDPLG 596

Qy   132 AKTVVT--------APEAIRGTRVQG--------------------TWYQKTVQLP 159
            ::::                ::::||                   :::::::::
Db   597 VSAVETRGDYKRFDDTYVANGDGIYIPSGWTGTMPNGDVIKPGVSFLDIRSFYKD---P 653

Qy   160 AGTKYVAFRHFGCTDFFWINLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDG 219
```

654 NWSKVQTFLDGGAEPQFRTHRFWAQTAVAGALADYARLFDDGT------------TTPDTTAP - 704
```
            || |            :: :             :::: |||||||
```

```
Qy    50 SFAGHNSAICVSSASYINFEGPQNPDNLVTPELSL------------PGGGTLTFWV---- 95
Db   205 QGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDV-TGATKVKYYAVNDG 278
```

705 ---TVPTG---LQAGVVTSEATISWT--ASTDDTRVTGYDVYRGATKV----------- 745
```
Db   206 AVANESTAVWLDRIGAI--EGNDSP----TTGSMGLRDHLEEAVRQSGGDPLTIQVVIYN 259
```

```
Qy    96 ----CAQDAN------YASEHYAVYASSTGN--DASNFANALLEEVLTAKTVVT 137
Db   279 FPGDHYAVMISKTGTNAGD---FTVVFEETPNGINKGGARFGLSTEA----DGAKPQSVWI 332
```

```
Db   260 LPGRDCAALASNGELGPDELDRYKSEYIDPIADIMWDFADYENLRIVAIIEIDSLPNLVT 319
```

746 --GTATTSFTDTGLTASTAYAYTVRAFDAAGNVSAPSAALTVTTKATFSDTTAP------- 798
```
```

```
Qy   138 APEAIRGTR-------VQGTWY--QKTVQLPAGTKYVAFRHFGCTDFFWINLD--- 181
Db   333 ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTTPDYT--------YTVYRDGT 385
```

```
Db   320 NVGGNGGTELCAYMKQNGGYVNGVGYALRKLGEIPNVYNYIDAAHHG-----WIGWDSNF 374
```

799 ---SVPAITS------------SSSTANSV---------TIGWSAS-TDNAGGSGLAGYDVYRGAT 839
```
```

386 KIKEGLTETFEEDGV-ATGNHEYCVEVKYTAG--VSPKVCVNVT 427
```
Qy   182 --EVEI-----KANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDW 232
```

```
Db   840 RVAQ-TTALFTDTGLTASTAYEYTVRARDVAGNVSAPSTAVSVT 883
```
Db   375 GPSVDIFYBAANASGSTVDYVHGFISNTANYSATVEPYLDVNGTVNGQLIRQSKWVDW 432
```

RESULT 15

C86822

hypothetical protein yqbK [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: C86822
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrl
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: C86822
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1649 <STO>
A;Cross-references: GB:AE005176; PIDN:AAK05677.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yqbK

Query Match   5.1%;   Score 121;   Pred. No. 3;   DB 2;   Length 1649;
Best Local Similarity   20.6%;   Pred. No. 3;
Matches 103;   Conservative 56;   Mismatches 182;   Indels 160;   Gaps 26;

```
Qy    46 PGGTSFAGHNSAICVSSASYINFEGPQN------------PDNY-------LVTPELSLPGG 88
            ||||| |||||| ||||||||||||||              :::
Db    65 PSGTTFNGNS---VSQPNRTPL--PQNSVEDAIMPNSFGWQQTIANLLTLPTNQVTAVQN 118
```

```
Qy    89 GTLTFWVCAQDANYASEHYAVYASSTGNDASNF------ANALLEEVLTAKTVVTAPEA-- 141
            || :: :: ::::::  :::::::  :::
Db   119 GTIIGSYPTTDAGVQQALYAMY-SAINASTSDFGLYYGANTTLSNSL-LKAVVSNPSATN 176
```

```
Qy   142 -----IRGTRVQGTWYQKTVQL--------PAGTKYVAFR----HFGCTDFFWINLDEVE 184
                  :::::::::                :::::::::
Db   177 MTFSTLKGHAKSLTWISNPADLLTSSNSSQPSGTNYTGTLPSNVIFGVPTIF----RNVT 232
```

```
Qy   185 IKANGKRADFTETFESSTHGEAPAEWTT------------------IDADGDGQGWL 223
            :::                              :::::::::
Db   233 VAASGDNIYAQGNAFATTNG----------SWITGAPNIYGGTDNSDISGNTNLYIGATGSIAGW- 287
```

```
Qy   224 CLSSGQLDWLTAHGGTNV----------------------VASFSWNGMALNPDNYLISKDVTG--ATKVK 270
```

```
Db   288 NIYGGNASAATISGNTHVTIAQSSTINSVTGGSASGTTISGNTNL---DISGAIASQIT 344
```

```
Qy   271 YYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNG----------INKGGARFG-----LST 320
```

```
Db   345 NIYGAGIG-------TSNSPVVNGNVTTYVNSTNGGARYQLYQGGTVYGNISGSIYN 395
```

```
Qy   321 EADGAKPQSVWIERTVDL---------PA---------------GTKYVAFRHYNCSDLNVILL 360
```

```
Db   396 TLSGA---GGWTGATSNINGAGGPASTFNGGSFQGNIGTSGAGNVISNSYNTSSFT---T 449
```

```
Qy   361 DDIQFTMGGSPTTPDYTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVEVKYTAGVSP 420
```

```
Db    450 GQALFTGGNAGTGASYAQA----------TNSTTAAQGILYANITNVIKSAFTTGTAG 497
Qy    421 KVC-------VNVTINPTQF 433
            :: |:|:|:
Db    498 AVYGIVGGNGHDSLKISPSQW 518
```

Search completed: May 18, 2004, 11:47:52
Job time : 13.4621 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:35:14 ; Search time 8.0235 Seconds
                                        (without alignments)
                                        2848.981 Million cell updates/sec

Title: US-08-570-311-22
Perfect score: 2388
Sequence: 1 GTPNPNPNPGTTTLSESF.........PKVCVNVTINPTQFNPVQNL 439

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SwissProt_42:*

                Pred. No. is the number of results predicted by chance to have a
                score greater than or equal to the score of the result being printed,
                and is derived by analysis of the total score distribution.

                                SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2388	100.0	2628	1	HGA2_PORGI	Q51845 porphyromon
2	2235	97.8	2164	1	HGA1_PORGI	P59915 porphyromon
3	1431	59.9	989	1	PRTH_PORGI	P46071 porphyromon
4	251.5	10.5	991	1	CPG1_PORGI	P28784 porphyromon
5	132.5	5.5	1256	1	ATL_STAAU	P52081 staphylococc
6	122.5	5.1	1090	1	GUXB_CELFI	P50899 cellulomona
7	120.5	5.0	2660	1	YEEJ_ECO57	Q8X8V7 escherichia
8	117.5	4.9	331	1	PME_ASPAC	Q12535 aspergillus
9	117.5	4.9	607	1	YK28_ARCFU	O28251 archaeoglob
10	117	4.9	1122	1	ADP1_MYCGA	Q49379 mycoplasma
11	113	4.7	1122	1	ADP2_MYCGA	Q9REM8 mycoplasma
12	112.5	4.7	1260	1	ALS1_CANAL	P46590 candida alb
13	112	4.7	872	1	GUXA_CELFI	P50401 cellulomona
14	111	4.6	465	1	SLAP_LACBR	Q05044 lactobacill
15	110	4.6	987	1	YD94_METJA	Q58789 methanococc
16	109.5	4.6	1385	1	C5AA_BACUD	Q45760 bacillus th
17	107.5	4.5	1034	1	ITAV_CHICK	P26008 gallus gall
18	107.5	4.5	1200	1	HYAL_STRPU	O76536 strongyloce
19	107.5	4.5	1953	1	BIGA_SALTY	P25927 salmonella
20	107	4.5	435	1	AM3D_ORYSA	P27933 oryza sativ
21	105.5	4.4	282	1	PRTA_ASPNG	P24665 aspergillus
22	105.5	4.4	699	1	CHI1_BACCI	P20533 bacillus ci
23	105.5	4.4	1609	1	FIG2_YEAST	P25653 saccharomyc
24	105	4.4	524	1	CHID_BACCI	P27050 bacillus ci
25	105	4.4	2812	1	ZAN_HUMAN	Q9Y493 homo sapien
26	104.5	4.4	551	1	AMYB_THETU	P19584 thermoanaer
27	104.5	4.4	1034	1	BGAL_BACME	O52847 bacillus me
28	104	4.4	721	1	OGP_MOUSE	Q62010 mus musculu
29	104	4.4	1443	1	NEO1_CHICK	Q90610 gallus gall
30	103	4.3	548	1	THER_BACST	P06874 bacillus st
31	102.5	4.3	527	1	NPRE_BACBR	P43263 bacillus br
32	102.5	4.3	1220	1	C5AC_BACTU	P56955 bacillus th
33	102	4.3	1044	1	ITA8_CHICK	P26009 gallus gall
34	102	4.3	1045	1	GUNB_CELFI	P26225 cellulomona
35	102	4.3	1289	1	C5AB_BACUD	Q45753 bacillus th
36	101.5	4.3	837	1	XYNZ_CLOTM	P10478 clostridium
37	101.5	4.3	2124	1	PGCA_RAT	P07897 rattus norv
38	101.5	4.3	2358	1	YEEJ_ECOLI	P76347 escherichia
39	101	4.2	439	1	SLAP_LACHE	P38059 lactobacill
40	101	4.2	827	1	XANP_XANS2	Q60106 xanthomonas
41	100.5	4.2	1377	1	NEO1_RAT	P97603 rattus norv
42	100.5	4.2	2249	1	OMPA_RICRI	P15921 rickettsia
43	100	4.2	623	1	TSPE_BPSFV	Q9XIP3 bacteriopha
44	100	4.2	762	1	E13B_TRIHA	P53626 trichoderma
45	100	4.2	1157	1	C8AA_BACUK	Q45704 bacillus th

                                ALIGNMENTS

RESULT 1
HGA2_PORGI
ID   HGA2_PORGI   STANDARD;   PRT;   2628 AA.
AC   Q51845;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97047672; PubMed=8926061;
RA   Han N., Whitlock J., Progulske-Fox A.;
RT   "The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381
RT   contains four large, contiguous, direct repeats.";
RL   Infect. Immun. 64:4000-4007(1996).
CC   -!- FUNCTION: Agglutinates erythrocytes.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U41807; AAB17128.1; -.
DR   PIR; T28651; T28651.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT   SIGNAL   1   24   POTENTIAL.
FT   CHAIN   25   2628   HEMAGGLUTININ A.
FT   DOMAIN   25   539   PEPTIDASE C25-LIKE 1.
FT   DOMAIN   540   995   PEPTIDASE C25-LIKE 2.
FT   DOMAIN   996   1451   PEPTIDASE C25-LIKE 3.
FT   DOMAIN   1452   1907   PEPTIDASE C25-LIKE 4.
FT   DOMAIN   2074   2628   PEPTIDASE C25-LIKE 5.
SQ   SEQUENCE   2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;

Query Match   100.0%; Score 2388; DB 1; Length 2628;
Best Local Similarity   100.0%; Pred. No. 1.6e-160;
Matches 439; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 GTPNPNPNPDGTTTLSESFENGIPASWKTIDADGDNNWTTTPPEGGTSFAGHNSAICV 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1862 GTPNPNPNPDGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 1921

Qy   61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1922 SSASYINFEGPQNPDNYLVTPELSLPDGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1981

```
Qy   121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1982 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 2041

Qy   181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  2042 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 2101

Qy   241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2102 VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV 2161

Qy   301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2162 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 2221

Qy   361 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2222 DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 2281

Qy   421 KVCVNVTINPTQFNPVQNL  439
             |||||||||||||||||||
Db  2282 KVCVNVTINPTQFNPVQNL 2300
```

RESULT 2
HGA1_PORGI
ID   HGA1_PORGI   STANDARD;   PRT;   2164 AA.
AC   P59915;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hemagglutinin A precursor.
GN   HAGA OR PG1837.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=22829867; PubMed=12994112;
RA   Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA   Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
RA   Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
RA   Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA   Dewhirst F.E., Fraser C.M.;
RT   "Complete genome sequence of the oral pathogenic bacterium
RT   Porphyromonas gingivalis strain W83.";
RL   J. Bacteriol. 185:5591-5601(2003).
CC   -!- FUNCTION: Agglutinates erythrocytes (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AE017178; AAQ66831.1; ALT_INIT.
DR   TIGR; PG1837; -.
KW   Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
KW   Complete proteome.
FT   SIGNAL   1   25   POTENTIAL.
FT   CHAIN   26   2164   HEMAGGLUTININ A.
FT   DOMAIN   26   539   PEPTIDASE C25-LIKE 1.
FT   DOMAIN   540   991   PEPTIDASE C25-LIKE 2.
FT   DOMAIN   992   1443   PEPTIDASE C25-LIKE 3.
SQ   SEQUENCE 2164 AA; 233387 MW; 6DFAB22832586C63 CRC64;

Query Match   97.8%;   Score 2335;   DB 1;   Length 2164;

Best Local Similarity   98.4%;   Pred. No. 6.9e-157;
Matches   430;   Conservative   2;   Mismatches   5;   Indels   0;   Gaps   0;

```
Qy     3 PNPNPNPNPGTTLSESFENGIPASWKTIPADGDGNNWTTTPPPGGTSFAGHNSAICVSS   62
            ||| ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db  1400 PNGTPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS 1459

Qy    63 ASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFA  122
            |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  1460 ASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFA 1519

Qy   123 NALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDE  182
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db  1520 NALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDD 1579

Qy   183 VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV  242
            |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  1580 VEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTNVV 1639

Qy   243 ASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVF  302
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1640 ASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVF 1699

Qy   303 EETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDD  362
            ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db  1700 EETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDD 1759

Qy   363 IQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKV  422
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1760 IQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKE 1819

Qy   423 CVNVTINPTQFNPVQNL  439
            |||||||||||||||||
Db  1820 CVNVTINPTQFNPVQNL 1836
```

RESULT 3
PRTH_PORGI
ID   PRTH_PORGI   STANDARD;   PRT;   989 AA.
AC   P46071;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protease prtH (EC 3.4.22.-).
GN   PRTH.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=95012612; PubMed=7927685;
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RT   "Cloning and characterization of a new protease gene (prtH) from
RT   Porphyromonas gingivalis.";
RL   Infect. Immun. 62:4279-4286(1994).
RN   [2]
RP   ERRATUM.
RA   Fletcher H.M., Schenkein H.A., Macrina F.L.;
RL   Infect. Immun. 62:5707-5707(1994).
CC   -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
CC   P.GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
CC   IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
CC   AND IS A VIRULENCE FACTOR.
CC   -!- SUBCELLULAR LOCATION: In membrane vesicles.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial

```
RL   J. Biol. Chem. 267:18896-18901(1992).
CC   -!- FUNCTION: Thiol protease which is believed to participate in
CC       intracellular degradation and turnover of proteins. Its
CC       proteolytic activity is a major factor in both periodontal tissue
CC       destruction and in bacterial host defense mechanisms. Activates
CC       complement C3 and C5.
CC   -!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
CC       immunoglobulins, with a preference for Arg in P1, and hydrophobic
CC       residues in P2 and P3.
CC   -!- ENZYME REGULATION: Requires cysteine for activation and Ca(2+)
CC       and/or Mg(2+) for stabilization. It is stimulated by glycine-
CC       containing dipeptides. It is resistant to inhibition by proteinase
CC       inhibitors in human plasma.
CC   -!- SIMILARITY: Belongs to peptidase family C25.
CC
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
DR   EMBL; D26470; BAA05484.1; -.
DR   PIR; I40229; I40229.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
KW   Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
FT   SIGNAL        1     24
FT   PROPEP       25    227
FT   CHAIN       228    991       GINGIPAIN R1.
FT   ACT_SITE    438    438       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    471    471       NUCLEOPHILE (BY SIMILARITY).
FT   CONFLICT    264    265       RT -> TK (IN REF. 2).
SQ   SEQUENCE   991 AA;  108782 MW;  03EE3F43CEBB2544 CRC64;

Query Match             10.5%;  Score 251.5;  DB 1;  Length 991;
Best Local Similarity   25.9%;  Pred. No. 4.1e-10;
Matches  120;  Conservative  50;  Mismatches  171;  Indels 123;  Gaps  21;

Qy    8 NPNPGTTTLSESFENGIPASWKTIDADGDG--NNWT----------TTPPP  46
          ||| |                                         :|||
Db  529 HPNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFGDPSLLVRTLVPTKMQVTAPAQ 588

Qy   47 GGTSFAG------HNSAICVSSASYINFEGPQNPDNYLVTPELS-LPGGGTLTFWVCAQD  99
           ||::          : :|  ::  |: |:  :   |  :  :  || : | :
Db  589 INLTDASVNVSCDYNGAIATISANGKMF-GSAVVENGTATINLTGLTNESLTLTLTV----- 643

Qy  100 ANYASEHYAVVASSTG--NDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQ 157
          :|  :           :|: : |                                    |
Db  644 VGYNKETVIKTINTNGEPNFYQPVSN-------LTATT----------QGQKVTLKWDAPSTK 689

Qy  158 LPAGTKYVAFRHFGCTDFFWINL-DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDAD 216
          || |  |  | : |  :   :|   :|:   :  | :|   |  |    || :|||
Db  690 TNATTN-TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWNDGS--GYQILLDAD 746

Qy  217 GDGQGWLCLSSGQLDWLTAHGGTNVVASFSWN------GMALNPDNYLISKDVTGATKVKY 271
          ||  |        |   | ||                :    |  |  :  ||| |  |
Db  747 HDQYGQVIPSDTHTLWPNCSVPANLFAPFEYTVPENADPSCSPTMIM--DGTASVNI--- 802

Qy  272 YYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINKGGARFGLSTEADGAKPQS-- 329
          : |                                              ||
Db  803 -------------DFAI---------AAPQANA 818

Qy  330 -VWI------ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTVTVY 381
            :|       :::| ||||  |   : |    :
Db  819 KIWIAGQGPFKEDDVVFEAGKKY-----HFLMKKMGSGDGTELTISEGGG---SDYTYTVY 871
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; L27483; AAA51298.1; -.
DR   HSSP; P23882; 1FMT.
DR   MEROPS; C25.001; -.
DR   InterPro; IPR002376; formyl transf.
DR   Pfam; PF00551; formyl transf; 1.
KW   Hydrolase; Thiol protease; Repeat; Virulence.
FT   REPEAT      270    323
FT   REPEAT      528    581
SQ   SEQUENCE   989 AA;  110238 MW;  FA85F8A3AC8944C CRC64;

Query Match             59.9%;  Score 1431;  DB 1;  Length 989;
Best Local Similarity   88.0%;  Pred. No. 1.9e-93;
Matches  271;  Conservative   8;  Mismatches  21;  Indels   8;  Gaps   3;

Qy  136 VTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFG----CTDFFWINLDEVEIKANGKR 191
          |||||||||||||||||||||:|||||||||||||    |:||||||||||||||||||
Db   34 VRSPEAIRG-RIQGTWRQKTVDLPAGTE:CCFPSLPKAPICST--STMRLRSKTNAKR  89

Qy  192 ADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMA 251
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   90 ADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVSSFSWNGMA 149

Qy  252 LNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINK 311
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  150 LNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVVFEETPNGINK 209

Qy  312 GGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP 371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  210 GGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSP 269

Qy  372 TPFDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTINPT 431
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||: |:|
Db  270 TPFDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKKCVNVTVNST 329

Qy  432 QFNPVQNL 439
          ||||:|||
Db  330 QFNPVKNL 337

RESULT 4
CPG1_PORGI
ID   CPG1_PORGI     STANDARD;     PRT;    991 AA.
AC   P28784; Q45168;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
DE   (RGP-1).
GN   RGPA OR RGP1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 228-290 AND 517-541.
RC   STRAIN=381;
RX   MEDLINE=95168884; PubMed=7864651;
RA   Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
RA   Ikehara Y.;
RT   "Structural characterization of arginingipain, a novel
RT   arginine-specific cysteine proteinase as a major periodontal
RT   pathogenic factor from Porphyromonas gingivalis.";
RL   Arch. Biochem. Biophys. 316:917-925(1995).
RN   [2]
RP   SEQUENCE OF 228-270.
RC   STRAIN=HG66;
RX   MEDLINE=92406812; PubMed=1527017;
RA   Chen Z., Potempa J., Polanowski A., Wikstrom M., Travis J.;
RT   "Purification and characterization of a 50-kDa cysteine proteinase
RT   (gingipain) from Porphyromonas gingivalis.";
```

```
Qy      382  RDGTKIKEGLTETTFEEDGVATGNHEYCVEVKKYTAGVSPKVCVN 425
             ||||||||||||||||  |:: .:|:::|||||||  |||||:|||:
Db      872  RDGTKIKEGLTETTYRDAGMSAQSHEYCVEVKKYAAGVSPKVCVD 915

RESULT 5
ATL_STAAU
ID   ATL_STAAU      STANDARD;      PRT;   1256 AA.
AC   P52081;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Bifunctional autolysin precursor [Includes: N-acetylmuramoyl-l-alanine
DE   amidase (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
DE   acetylglucosamidase (EC 3.2.1.96)].
GN   ATL.
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
RC   STRAIN=RN450;
RX   MEDLINE=95116542; PubMed=7816834;
RA   Oshida T., Sugai M., Komatsuzawa H., Hong Y.-M., Suginaka H.,
RA   Tomasz A.;
RT   "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
RT   alanine amidase domain and an endo-beta-N-acetylglucosamidase
RT   domain: cloning, sequence analysis, and characterization.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 8325-4;
RA   Foster S.J.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOBIOSYL UNIT IN
CC       HIGH-MANNOSE GLYCOPEPTIDES AND GLYCOPROTEINS CONTAINING THE
CC       -[(MAN)5(GLCNAC)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
CC       RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
CC       OLIGOSACCHARIDE IS RELEASED INTACT.
CC   -!- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
CC       residues and L-amino acid residues in certain bacterial cell-wall
CC       glycopeptides.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC       unit in high-mannose glycopeptides and glycoproteins containing
CC       the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC       residue remains attached to the protein; the rest of the
CC       oligosaccharide is released intact.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
CC       EXTRACELLULAR LYTIC ENZYMES.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
CC       ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
CC       GLYCOSYL HYDROLASES.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; D17366; BAA04185.1; -.
DR   EMBL; L41499; AAA99982.1; -.
DR   InterPro; IPR002502; Amidase_2.
DR   InterPro; IPR002901; Amidase_4.
DR   Pfam; PF01510; Amidase_2; 1.
DR   Pfam; PF01832; Amidase_4; 1.
DR   SMART; SM00644; Ami_2; 1.
DR   SMART; SM00047; LYZ2; 1.
KW   Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
FT   SIGNAL        1     29       POTENTIAL.
```

```
FT   CHAIN        30   1256       BIFUNCTIONAL AUTOLYSIN.
FT   DOMAIN      199    775       N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
FT   DOMAIN      776   1256       ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
FT   REPEAT      425    589       1.
FT   REPEAT      596    758       2.
FT   REPEAT      770    932       3.
SQ   SEQUENCE   1256 AA;  137384 MW;  2BB76CAA292FDD20 CRC64;

Query Match             5.5%;  Score 132.5; DB 1; Length 1256;
Best Local Similarity  21.9%;  Pred. No. 0.15;
Matches 103;  Conservative  47; Mismatches 167;  Indels 153;  Gaps  25;

Qy    3  PNPNPNPNPGTT---LSESFENGIPASWKTIDADGDG---NNWTTTPPP--- 46
         |||  |::| ||    :|    :::|:|: |:|:|
Db  594  PTPTPTTPKPSTPTTNNKLTVSSLNGV----AQINAKNNGLFTTVYDKTGKPTKBVQKTFA 649

Qy   47  ------GGTSF--AGHNSAICV---SSASYINPEGPGN-PDNYLVTPELSL--- 85
               :||::|  ::|| ::|     :::|::|  || :
Db  650  VTKEASLGGNKFYLVKDYNSPTLIGWVKQGDVIYNNAKSPVNVMQTYTVKPGTKLYSVPW 709

Qy   86  ------PGGGTLTFWVCAQDANYASEHYAVVASSTGNDASNFAN------ALLE 127
               :.: :||    :  :|||::|:|||::|:::       .::
Db  710  GTYKQEAGAVSGTGNQTFKATKQQQIDKS----IYLFGTVNGKSGWVSKAYLAVPAAPKK 765

Qy  128  EVLTAKT----VVTAPEAIR-----------GTRVQGTWYQKTVQLPAGTKYVAFRHFG 171
         |  ::|     || |||:|           ||:||  |:| |:|  ||||:|
Db  766  AVAQPKTAVKAYTVTKPQTTQTVSKIAQVKPNNTGIRASVVEKTAK--NGAKY-ADRTFY 822

Qy  172  CT----------------DFFWINLDEVEIKANGKRADFTE-TFESSTHGEAPAE 209
         |:                |::| :||:  ||:| |  | :  |:|:::| |:|
Db  823  VTKERAHGNETYVLLNNTSHNIPLGWFNVKDLNVQNLGKEVKTTQKYTVNKSNNGLSMVP 882

Qy  210  WTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWN-----GMALNPDNYLISKDV 263
         |::|| :: || | :     :|  | :| || :|:|       |  :|:|||:|:.:
Db  883  WGT-------KNQVILTGNNI---AQGTFNATKQVSVGKDVYLYGTINNRTGWVNAKDL 931

Qy  264  TGATKVK----------YYYAVNDGFPGDHYAVMISKTGT------NAGDFTVVFEETPN 307
         || |:|           ||:|| |||:|||::|::||         :|||||:|:|:||
Db  932  TAPTAVKPTTSAAKDYNYTVIKNG-NGYYVTPNSDTAKYSLKAFNEQPFAVVKEQVIN 990

Qy  308  G------INKGGARFGLSTEADGAK-----PQSVWIERTVDLPAGTKY 344
                ::|||:|::| : |||:        || |:|| :|| |||
Db  991  GQTWYYGKLSNGKLAWIKST--DLAKELIKYNQTGWTLNQVAQIQAGLQY 1038
```

```
RESULT 6
GUXB_CELFI
ID   GUXB_CELFI     STANDARD;      PRT;   1090 AA.
AC   P50899;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Exoglucanase B precursor (EC 3.2.1.91) (Exocellobiohydrolase B)
DE   (1,4-beta-cellobiohydrolase B) (CBP120).
GN   CBHB OR CENE.
OS   Cellulomonas fimi.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Micrococcineae; Cellulomonadaceae; Cellulomonas.
OX   NCBI_TaxID=1708;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
RC   STRAIN=ATCC 484;
RX   MEDLINE=96003898; PubMed=7575482;
RA   Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT   "Cellobiohydrolase B, a second exo-cellobiohydrolase from the
RT   cellulolytic bacterium Cellulomonas fimi.";
RL   Biochem. J. 311:67-74(1995).
RN   [2]
RP   SEQUENCE OF 54-75.
RX   MEDLINE=93209933; PubMed=8458833;
RA   Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT   "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RT   D (CenD), a family A beta-1,4-glucanase.";
```

```
Qy  132  AKTVVT-------------------------APEAIRGTRVQG-----------------TWYQKTVQLP 159
                                                                    ::|:
Db  597  VSAVETRGDYKRFDDTYVANGDGIYIPSGWTGTMPNGDVIKPGVSFLDIRSFYKKD---P 653

Qy  160  AGTKYVAFRHFGCTDFFWINLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDG 219
               ::|:|
Db  654  NWSKVQTFLDGGAEPQFRYHRFWAQTAVAGALADYARLPDDGT------TPDTTAP- 704

Qy  220  QGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDV-TGATKVKYYYAVNDG 278
               ::|        |::
Db  705  ----TVPTG---LQAGVVTSEATISWT--ASTDDTRVTGYDVYRGATKV--------- 745

Qy  279  FPGDHYAVMISKTGTNAGD---FTVVFEETPNGINKGGARFGLSTEA---DGAKPQSVWI 332
               |:|      ::    :|:      ::|:
Db  746  --GTATTTSFTDTGLTASTAYAYTVRAFDAAGNVSAPSAALTVTTKATPSDTTAP---- 798

Qy  333  ERTVDLPAGTKYVAFRHVNCSDLNYILLDDIQFTMGGSPTPTDYT------YTVYRDGT 385
               ::||      :|       :
Db  799  ---SVPAITS------SSSTANSV----TIGWSAS-TDNAGGSGLAGYDVYRGAT 839

Qy  386  KIKEGLTETTFEEDGV-ATGNHEYCVEVKYTAG--VSPKVCVNVT 427
               :||:  ::||           ||:|:
Db  840  RVAQ-TTALTFTDTGLTASTAYEYTVRARDVAGNVSAPSTAVSVT 883
```

```
RESULT 7
YEEJ_ECO57
ID   YEEJ_ECO57    STANDARD;    PRT;   2660 AA.
AC   Q8X8V7; Q8X2B9; Q8X2C0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein yeeJ.
GN   Z3315 OR ECS2775/ECS2776.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunket G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- SIMILARITY: Contains 16 Big-1 domains.
CC   -!- SIMILARITY: Belongs to the intimin/invasin family.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 1315.
```

(Left column — first database entry)

```
RL   J. Bacteriol. 175:1910-1918(1993).
RN   [3]
RP   SEQUENCE OF 54-78.
RX   MEDLINE=9419770B; PubMed=8147863;
RA   Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.;
RA   Warren R.A.J., Miller R.C. Jr.;
RT   "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi
RT   CenE, a member of a new family of beta-1,4-glucanases.";
RL   Biochem. Biophys. Res. Commun. 199:1223-1228(1994)
CC   -!- FUNCTION: Hydrolyze cellohexaose to a mixture of cellotetraose,
CC       cellotriose and cellobiose, with only a trace of glucose. It
CC       hydrolyzed cellopentaose to cellotriose and cellobiose, and
CC       cellotetraose to cellobiose, but it did not hydrolyze cellotriose.
CC       Has also weak endoglucanase activity. Hydrolyzes glucosidic bonds
CC       with inversion of anomeric configuration.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC   -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC       domain.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL
CC       HYDROLASES).
CC   -----------------------------------------------------------------
```

```
DR   EMBL; L38827, AAB00822.1; -.
DR   PIR; S59077; S59077.
DR   HSSP; P07986; 1EXG.
DR   InterPro; IPR001919; Bac_celose-bind.
DR   InterPro; IPR008965; Cellul_bind.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR000556; Glyco_hydro_48.
DR   InterPro; IPR008928; Glyco_trans_6hp.
DR   Pfam; PF00553; CBM_2; 1.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF02011; Glyco_hydro_48; 1.
DR   PRINTS; PR00844; GLHYDRLASE48.
DR   ProDom; PD011903; Glyco_hydro_48; 1.
DR   SMART; SM00637; CBD_II; 1.
DR   SMART; SM00060; FN3; 3.
DR   PROSITE; PS00561; CBD_BACTERIAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT   SIGNAL         1       33
FT   PROPEP        34       53
FT   CHAIN         54     1090       EXOGLUCANASE B.
FT   DOMAIN        54      699       CATALYTIC (BY SIMILARITY).
FT   DOMAIN       700      785       FIBRONECTIN TYPE-III 1.
FT   DOMAIN       794      884       FIBRONECTIN TYPE-III 2.
FT   DOMAIN       891      978       FIBRONECTIN TYPE-III 3.
FT   DOMAIN       989     1090       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE     513      513       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID     990     1089       BY SIMILARITY.
SQ   SEQUENCE   1090 AA;  114829 MW;  046BB9D956F2F399 CRC64;
```

```
Query Match    5.1%;  Score 122.5;  DB 1;  Length 1090;
Best Local Similarity  21.9%;  Pred. No. 0.61;
Matches 102;  Conservative 47;  Mismatches 157;  Indels 159;  Gaps   25;

Qy    26  ASWKT--IDADGDGNNWTTPPPGGTSFAGHNSAICVSSASYINFEGPQNPDNYLVTPE 82
              ||||
Db   515  ASWKVPSELKWTGKPDTWNAAAPTG------NPGLTVEVTSY------------ 550

Qy    83  LSLPGGGTLTFWVCAQDANYASE--HYAVYASSTGNDAS-NFANALLEV------LT 131
              ||        :: ::|    |||: :|||:  ::  :||||
Db   551  ----------GQDVGVAADTARALLFYAAKSGDTASRDKAKALLDAIWANNQDPLG 596
```

CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; AE005423; AAG57041.1; -.
DR   EMBL; AP002559; BAB36198.1; ALT_FRAME.
DR   EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR   InterPro; IPR003344; Big_1.
DR   InterPro; IPR003535; Intimin.
DR   InterPro; IPR008964; Invasin_intimin.
DR   InterPro; IPR000601; PKD.
DR   Pfam; PF02369; Big_1; 16.
DR   PRINTS; PR01369; INTIMIN.
DR   SMART; SM00634; BID_1; 16.
DR   SMART; SM00089; PKD; 8.
KW   Hypothetical protein; Repeat; Complete proteome.
FT   DOMAIN   738   834   BIG-1 1.
FT   DOMAIN   840   929   BIG-1 2.
FT   DOMAIN   931   1033   BIG-1 3.
FT   DOMAIN   1042   1132   BIG-1 4.
FT   DOMAIN   1134   1236   BIG-1 5.
FT   DOMAIN   1245   1335   BIG-1 6.
FT   DOMAIN   1337   1439   BIG-1 7.
FT   DOMAIN   1448   1539   BIG-1 8.
FT   DOMAIN   1548   1652   BIG-1 9.
FT   DOMAIN   1653   1750   BIG-1 10.
FT   DOMAIN   1751   1855   BIG-1 11.
FT   DOMAIN   1856   1957   BIG-1 12.
FT   DOMAIN   1963   2056   BIG-1 13.
FT   DOMAIN   2065   2156   BIG-1 14.
FT   DOMAIN   2157   2252   BIG-1 15.
FT   DOMAIN   2254   2355   BIG-1 16.
SQ   SEQUENCE   2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match   5.0%; Score 120.5; DB 1; Length 2660;
Best Local Similarity 22.2%; Pred. No. 2.8;
Matches 105; Conservative 64; Mismatches 192; Indels 111; Gaps 26;

Qy   13 TTTLSESFE---NGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICVSSASYINFE   69
           |:|::| | |:| | : | ||||||:| | | |:   ::
Db   1470 TATVKDQFDNEVNNLPVTFSTA-----SSGLTLPGBSNT---NESGIAQATLAGVAF-   1519

Qy   70 GPQNPDNYLVTPELSLPLGGG---TLTF--WVCAQDANYASEHYAVYASSTGNDASNFAN   123
           : :   :: || ::   ::|| :|
Db   1520 GEQT----VTASLANNGASDNKTVHFIGDTAAAKIIELTPVPDSIIAGTPQNSSGGS---   1571

Qy   124 ALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEV   183
           : | | : : || || ||: | | || : |:|:| :: |||
Db   1572 ----VITATVVDNNGFPVKGVTVNFTSNAATAEMNGGQAVTNEQGKATVTYTNTRSSI   1626

Qy   184 EIKANGKRADFTETFESSTHGEAPAWTTIDADGDGQ-GWLCLSSGQLDWLTAHGGTNV-   241
           || :| | : : | | ||||||:| |:| |:|||| |||||||| |
Db   1627 B---SGARPD---TVEASLENGSSTLSTSINVNADASTAHLTLLQALFDTVSAGDTTNLY   1680

Qy   242 --VASFSWNG------MALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTG   292
            | | :||| || ||||| |:||| |||| |||:||||||||| ::||
Db   1681 IEVKDNYGNGVPQQEVTLSVSP---SEGVTPSNNAIY---TTNHDGNFYA---SFTA   1728

Qy   293 TNAGDFTVV-------FEETPNGI-NKGGARFGLSTEADGAKPQSWIERTVDLPAGTK   343
           || | | ||| :| |||||:|||||||||||:||||||||||||||||
Db   1729 TKAGVQVTATLENGDSMQQTVTYPVPNVANAEISLAASKDPV------IANNNDLTLTA   1782

Qy   344 YVAFRHYNC--SDLNYILLDDIQ--FTMGGSPTPDTYTYVYRDGTKI---KEGLTETT   395
           |:| | ||| |:| :| | :: ::||| |||:|::|| | ||| |:|
Db   1783 TVADTEGNAIANSEVTFTLPEDVRANFTLG----------DGGKVVTDTEGKAKVT   1828

Qy   396 FEEDGVATGNHEYCVEVKYTAGVSPKVCVN------------VTINPTQFNPVQN   438
           | | |:| | : :| | | || ||:|| ::|::|
Db   1829 LK--GTKAGAH--TVTASMAGGKSEQLVVNFIADTLTAQVNLNVTEDNFIAN   1876

RESULT 8
PME_ASPAC
ID   PME_ASPAC   STANDARD;   PRT;   331 AA.
AC   Q12535;

DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Pectinesterase precursor (EC 3.1.1.11) (Pectin methylesterase) (PE).
GN   PME1.
OS   Aspergillus aculeatus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5053;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KSM 510;
RX   MEDLINE=97079238; PubMed=8920970;
RA   Christgau S., Kofod L.V., Halkier T., Andersen L.N., Hockauf M.,
RA   Dorreich K., Dalboege H., Kauppinen S.;
RT   "Pectin methyl esterase from Aspergillus aculeatus: expression
RT   cloning in yeast and characterization of the recombinant enzyme.";
RL   Biochem. J 319:705-712(1996).
CC   -!- FUNCTION: Involved in maceration and soft-rotting of plant tissue.
CC   -!- CATALYTIC ACTIVITY: Pectin + N H(2)O = N methanol + pectate.
CC   -!- SIMILARITY: Belongs to the pectinesterase family.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U49378; AAB42153.1; -.
DR   InterPro; IPR000070; Pectinesterase.
DR   Pfam; PF01095; Pectinesterase; 1.
DR   PROSITE; PS00800; PECTINESTERASE_1; 1.
DR   PROSITE; PS00503; PECTINESTERASE_2; 1.
KW   Hydrolase; Aspartyl esterase; Cell wall; Signal.
FT   SIGNAL   1   17   BY SIMILARITY.
FT   CHAIN   18   331   PECTINESTERASE.
FT   ACT_SITE   162   162   BY SIMILARITY.
FT   ACT_SITE   183   183   BY SIMILARITY.
SQ   SEQUENCE   331 AA; 35681 MW; 1F1C1BF1E32174F CRC64;

Query Match   4.9%; Score 117.5; DB 1; Length 331;
Best Local Similarity 22.9%; Pred. No. 0.29;
Matches 70; Conservative 33; Mismatches 98; Indels 105; Gaps 15;

Qy   108 AVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGT-KYVA   166
          | |: |:| || | : ||: : || || ||::| :| ||:
Db   27 AIVVAKSGGDYTTIGDAI--DALSTSTDTQIFIE---EGT-YDEQVYLPAMTGKVII   79

Qy   167 FRHFGCTDFFWINL----DEVEIKANGKRADFTETFE------------SSTHGEAPAE   209
          : | |: ::| | || ||: || |: : | |::| |:|:
Db   80 YGQTENTDSYADNLVITHAISYEDAGESDLTATFRNKAVGSQVNLNIANTCGQACHQ   139

Qy   210 WTTIDADGDGQWL-CLSSGQLDWLTAHGGTNVV------------------------   242
          ||||||||||| |:|||||| ||||||| ||
Db   140 ALALSAWADQQGYYGCNFTGYQDTLLAQTGNQLYINSYIEGAVDFIFGQHARAWFQNVDI   199

Qy   243 ------ASFSWNGMALNPDN--YLISKDVTGA------TKVKYYYAVNDGFPGDHYAV   286
                 |||| | |:|||| |:| |||| | ||||::| |||: || |:
Db   200 RVVEGPTSASTANGRSSETDSYYVINKSTVAAKEGDDVAEGTYL---GRPWSEYA-   254

Qy   287 MISKTGNAGDFTVVFEET--PNGINKGGARFGLSTEADGAKPQSWIERTVDLPAGTKY   344
          |::|||| |||| | : ||| || |:| | | ::|| | |:|| | |
Db   255 -------RVVFQQTSMTNVINSLG---------------------WTEWSTSTP-NTEY   284

Qy   345 VAFRHY 350
          |:|:|
Db   285 VTFGEY 290

RESULT 9
YK28_ARCFU

```
Db    434  DIY--NITLVVENLGGGESPYVYVYDLIPK------NFSLTNGDNDWKDPQDRDGDWV 483
                 |.:         .::        |.|:
Qy    333  ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTV---------- 380
                 |:           ::         :::|.||
Db    484  NKSSMLAGGPETIT--------------------NIQL---SGYDTGYWRIRPINASADGDGA 524

Qy    381  YRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTINP 430
                 |||:|        :|:        |.|        :|||:
Db    525  YDDYTEIENNQTVIFYQ--IQGSEDYKLLDAFIVGIDPILSMNEQTSP 571
```

RESULT 10

```
ID   ADP1_MYCGA     STANDARD;          PRT;   1122 AA.
AC   Q49379; Q49437; Q53351;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Adhesin P1 precursor (Cytadhesin P1) (Attachment protein).
GN   GAPA OR MGC1.
OS   Mycoplasma gallisepticum.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2096;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S6;
RX   MEDLINE=96201559; PubMed=8613358;
RA   Keeler C.L. Jr., Hnatow L.L., Whetzel P.L., Dohms J.E.;
RT   "Cloning and characterization of a putative cytadhesin gene (mgc1)
RT   from Mycoplasma gallisepticum.";
RL   Infect. Immun. 64:1541-1547(1996).
RN   [2]
RP   SEQUENCE OF 1-12 FROM N.A.
RC   STRAIN=S6;
RA   Hnatow L.L., Keeler C.L. Jr., Tessmer L., Dohms J.E.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 378-570 FROM N.A.
RC   STRAIN=S6;
RX   MEDLINE=93371270; PubMed=8363503;
RA   Dohms J.E., Hnatow L.L., Whetzel P., Morgan R., Keeler C.L. Jr.;
RT   "Identification of the putative cytadhesin gene of Mycoplasma
RT   gallisepticum and its use as a DNA probe.";
RL   Avian Dis. 37:380-388(1993).
RN   [4]
RP   SEQUENCE OF 159-1122 FROM N.A.
RC   STRAIN=S6;
RA   Goh M.S., Geary S.J.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Could be involved in cytadherence.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the adhesin P1 family.
CC   -!- CAUTION: Ref 4 sequence differs from that shown due to a
CC       frameshift in position 159 to 213.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U34842; AAB02987.1; -.
DR   EMBL; U44804; AAC83385.1; ALT_FRAME.
DR   PIR; T18346; T18346.
KW   Cytadherence; Signal; Transmembrane.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1122       ADHESIN P1.
FT   TRANSMEM    997   1021       POTENTIAL.
FT   DOMAIN     1002   1009       POLY-ILE.
FT   CONFLICT    313    314       DM -> IW (IN REF. 4).
```

```
ID   YK28_ARCFU     STANDARD;          PRT;    607 AA.
AC   O28251;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AF2028.
GN   AF2028.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: TO M.JANNASCHII MJ1393 AND MJ1394.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE000963; AAB89227.1; -.
DR   PIR; C69503; C69503.
DR   TIGR; AF2028;
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     16     36       POTENTIAL.
FT   TRANSMEM    583    603       POTENTIAL.
SQ   SEQUENCE   607 AA;  66215 MW;  F46F116BF2AAFEE1 CRC64;
```

```
Query Match             4.9%;  Score 117.5;  DB 1;  Length 607;
Best Local Similarity  18.3%;  Pred. No. 0.64;
Matches   86;  Conservative   65;  Mismatches  172;  Indels 147;  Gaps   23;

Qy      9  PNPGTTTLSESFENGIPASWKTID-ADG--------------DGNNWTTPPPGGTSFAGHN  55
              |:  |.|::::        :||||| |                ||| ||
Db      1  PSSVTILTVNKNLNTSYDTNWKTLEFADGSSNPKQIISGTLDNNKWLTTT----------- 249

Qy     56  SAICVSSASYINFEGPQNPDNVLVTPELSLPGGGTLTFWVCAQDAN----------YASEH 106
              :    ||    |  |   |        |      |  |||||
Db    250  ------------FKVKGNYSTDEAFVS--RLVGFGEAYFQFGPINGNISGTKIIDVFAIGN 296

Qy    107  YAVYASSTGNDASN--FANALLEEVLTAKTVVTAPEAIRGT----RVQGTWYQKTVQLP 159
              :: ||        ::|      :|  |     |||           || |  |||
Db    297  ASIGVNKSGPDNSNQWTGNVTIKNTATGLTYIVKSVKVWATDRNYNEINGARYENTT---- 353

Qy    160  AGTKYVAFRHFGCTDFFWINLDEVEIKANGKRADFTETFESSTHGEAPAEWTTID--AD 216
              |        |       |   |     |        ::|      |   |||
Db    354  -------------VN---VQI---GRDESFTSKDLSFQYDKVPIIWGNVTFRLVE 389

Qy    217  GDGQGWL CLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVN 276
              |  |||      :| |      ::  |     |     | |::: ::|  ||
Db    390  DANYGW---GVGQ-DKITDGGNTYIIERI-----------YVIGSYLVKVTK-HVESAGN 433

Qy    277  DGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGGARFGLST--EADGAKPQ--SVWI 332
```

```
FT   CONFLICT       382    382       Y -> F (IN REF. 3).
FT   CONFLICT       501    501       I -> T (IN REF. 3).
FT   CONFLICT       568    568       R -> G (IN REF. 3).
FT   CONFLICT       570    570       T -> A (IN REF. 3).
FT   CONFLICT       693    695       DIL -> VYT (IN REF. 4).
SQ   SEQUENCE   1122 AA;  121305 MW;  155C34DA2D6C3C65 CRC64;

Query Match             4.9%;  Score 117;  DB 1;  Length 1122;
Best Local Similarity  19.5%;  Pred. No. 1.6;
  Matches   93;  Conservative   78;  Mismatches  195;  Indels 112;  Gaps  24;

Qy   42  TTPPPGGTSFAGHNSAICVSSASYINFEGPQN----------PDNYLVTPELSLPGGGT   90
                                                      :::  :
Db   437 TAEAPGNTKTVGYPYGILLSA---ISFDATRNGLALAPASLGQDVGYHFVPRLAVGGVSS 493

Qy   91  LTFWVCAQDANYASEHYAVYASSTGNDASNFAN------ALLEEVLTAKTVVTAPEAIRG 144
Db   494 -----PRGAN--GNIFLGSAITWGTNGGNFLDTKWHSPAVIEDAPTTFVTVNSSGVLQN   545

Qy   145 TRVQGTWYQKTVQLP--AGTKYVAFRHFGCTDFFWINLDEVEIK---ANGKRAD--FTET 197
Db   546 S---GSQQSTSTPMPNSNGNESIPYRWTNSYDYNSVRFAALISKPAGGNTKQVESLFTTA 602

Qy   198 FESSTHGEAPAEWTTID------ADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWN---  248
Db   603 LKLDTLNSLPNKFTQENNIFFSYAMLDGRQWSLGTRKDSTWLT----TNTINNFTYNTQQ 658

Qy   249 ------------GMALNPDNVLIS---------KDV--------TGATKVKYYYAVND---  277
Db   659 QLASTAAGENANPRNILNALTTAKGFDRRDIGNVDILYSNNTNKFTYYYQVGGAITTWPE  718

Qy   278 ---GFPGDHYAVMISKTGTNAGDFTVVFEETPNGINK-GGARFGLSTEADGAKPQSVWIE  333
Db   719 VQVNYKTSANITYYNLTRTDFGSTTPATQDANTVSSKLNGAYLSSTGDQQGWYNGSIYVK  778

Qy   334 RTVDLPAGTKYV--AFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGL  391
Db   779 KASFTPSSQGYTWQDFKGLTTTASNAVISN--WTKAG------YSIRPDDDTV-FSV     826

Qy   392 TETTFEEDGVATGN----HEYCVE-----VKYTAGVSPKVCVNVTINPTQF-NPVQN    438
Db   827 SKIPFEKEITAAVNVRSLDSYYVQLNGETSVNTVARVSPDSSA-LTLNPKRITNPLMN  883
```

```
RESULT 11
ADP2_MYCGA
ID   ADP2_MYCGA     STANDARD;     PRT;   1122 AA.
AC   Q9REM8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Adhesin P1 precursor (Cytadhesin P1) (Attachment protein) (Adherence
DE   protein A).
GN   GAPA OR MYCGA1800 OR MGA_0934.
OS   Mycoplasma gallisepticum.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2096;
```
```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R(low);
RX   MEDLINE=22830409; PubMed=12949158;
RA   Papazisi L., Gorton T.S., Kutish G., Markham P.F., Browning G.F.,
RA   Nguyen D.K., Swartzell S., Madan A., Mahairas G., Geary S.J.;
RT   "The complete genome sequence of the avian pathogen Mycoplasma
RT   gallisepticum strain R(low).";
RL   Microbiology 149:2307-2316(2003).
RN   [2]
RP   SEQUENCE OF 94-1122 FROM N.A.
RC   STRAIN=R(low);
RX   MEDLINE=20536401; PubMed=11083776;
RA   Papazisi L., Troy K.E., Gorton T.S., Liao X., Geary S.J.;
RT   "Analysis of cytadherence-deficient, GapA-negative Mycoplasma
RT   gallisepticum strain R.";
RL   Infect. Immun. 68:6643-6649(2000).
```

```
CC   -!- FUNCTION: Could be involved in cytadherence (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the adhesin P1 family.
```

```
DR   EMBL; AE016967; AAP56530.1; ALT_INIT.
DR   EMBL; AF214004; AAF25381.1; -.
KW   Cytadherence; Signal; Transmembrane; Complete proteome.
FT   SIGNAL          1     26       POTENTIAL.
FT   CHAIN          27   1122       ADHESIN P1.
FT   TRANSMEM     1001   1021       POTENTIAL.
FT   DOMAIN       1002   1009       POLY-ILE.
FT   CONFLICT      337    337       F -> C (IN REF. 2).
FT   CONFLICT      645    645       T -> A (IN REF. 2).
FT   CONFLICT      826    826       V -> A (IN REF. 2).
SQ   SEQUENCE   1122 AA;  121387 MW;  E57D27A7FBD360AD CRC64;

Query Match             4.7%;  Score 113;  DB 1;  Length 1122;
Best Local Similarity  19.5%;  Pred. No. 3;
  Matches   93;  Conservative   77;  Mismatches  196;  Indels 112;  Gaps  24;

Qy   42  TTPPPGGTSFAGHNSAICVSSASYINFEGPQN----------PDNYLVTPELSLPGGGT   90
Db   437 TAEAPGNTKTVGYPYGILLSA---ISFDATRNGLALAPASLGQDVGYHFVPRLAVGGVSS 493

Qy   91  LTFWVCAQDANYASEHYAVYASSTGNDASNFAN------ALLEEVLTAKTVVTAPEAIRG 144
Db   494 -----PRGAN--GNIFLGSAITWGTNGGNFLDTKWHSPAVIEDAPTTFVTVNSSGALQN   545

Qy   145 TRVQGTWYQKTVQLP--AGTKYVAFRHFGCTDFFWINLDEVEIK---ANGKRAD--FTET 197
Db   546 S---GNPQPTSTPMPNSNGNESIPYRWTNSYDYNSVRFAALISKPAGGNTKQVESLFTTA 602

Qy   198 FESSTHGEAPAEWTTID------ADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWN---  248
Db   603 LKLDTLNSLPNKFTQENNIFFSYAMLDGRQWSLGTRKDSTWLT----TNTINNFTYNTQQ 658

Qy   249 ------------GMALNPDNVLIS---------KDV--------TGATKVKYYYAVND---  277
Db   659 QLASTVAGENANPRNILNALTTAKGFDRRDIGNVVTYSNNTNKFTYYYQVGGAITTWPE  718

Qy   278 ---GFPGDHYAVMISKTGTNAGDFTVVFEETPNGINK-GGARFGLSTEADGAKPQSVWIE  333
Db   719 VQVNYKTSANITYYNLTRTDFGSTTPATQDANTVSSKLNGAYLSSTGDQQGWYNGSIYVK  778

Qy   334 RTVDLPAGTKYV--AFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGL  391
Db   779 KASFTPSSQGYTWQDFKGLTTTASNAVISN--WTKAG------YSIRPDDDTV-FNV     826

Qy   392 TETTFEEDGVATGN----HEYCVE-----VKYTAGVSPKVCVNVTINPTQF-NPVQN    438
Db   827 SKIPFEKEITAAVNVRSLDSYYVQLNGETSVNTVARVSPDSSA-LALNPNRITNPLMN  883
```

```
RESULT 12
ALS1_CANAL
ID   ALS1_CANAL     STANDARD;     PRT;   1260 AA.
AC   P46590;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Agglutinin-like protein 1 precursor.
GN   ALS1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
```

```
Db  432 PSPNPTVSTEYWSQSF-----ATTTVTAPPGGTDTVIIREPPNHTVTTEYWSQSFAT 486
            |:||| ||          |:          :|:        ||
Qy   41 --TTPPGGTSPAGHNSAICVSSASYINFEGPQNP-------DNYLVTPELSLPGGG 89
             |                                         : |
Db  487 TTTVTAPPGGT-----DSVI-----IREPNPTVTTEYWSQSFATTTVTVTAPPGG 532
            |:|| ||                    |
Qy   90 TLTFWV------CAQDANYASEHYAVYASSTGNDASNFANALLEVLTAKTVVTAPEAIR 143
                                                               ||||
Db  533 TDSVIIREPPNPTVTTEYWSQSYA--- 565
             |:| |||||                |
Qy  144 GTRVQGTWYQKTVQLPAGTKYVAFR----HFGCTDFFWINLDEVEIKANGKRADFTETFE 199
                                                                   ::
Db  566 ----------------PGGTDSVIIREPPNHTVTTEYW--- 593
                                ::
Qy  200 SSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLI 259
                                                           ::|:|
Db  594 TTTTVTAPP-----GGTDTV--- 618
             ::|                   |||||||||:|
Qy  260 SKDVTGATKVKYYAVNDGFPGDHYAVMLSKTGTNAGDFTVVFEETPN 307
                                                          :|: |||||
Db  619 ----TTTEYW------SQSFATTTVTGPPSGTDTVIIREPPN 651
```

RESULT 13
GUXA_CELFI
```
ID   GUXA_CELFI     STANDARD;      PRT;   872 AA.
AC   P50401;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Exoglucanase A precursor (EC 3.2.1.91) (Exocellobiohydrolase A)
DE   (1,4-beta-cellobiohydrolase A) (CBP95).
GN   CBHA.
OS   Cellulomonas fimi.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Micrococcineae; Cellulomonadaceae; Cellulomonas.
OX   NCBI_TaxID=1708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 484;
RX   MEDLINE=94344030; PubMed=8065260;
RA   Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,
RA   Miller R.C. Jr.;
RT   "Cellobiohydrolase A (CbhA) from the cellulolytic bacterium
RT   Cellulomonas fimi is a beta-1,4-exocellobiohydrolase analogous to
RT   Trichoderma reesei CBH II.";
RL   Mol. Microbiol. 12:413-422(1994).
RN   [2]
RP   SEQUENCE OF 41-58.
RX   MEDLINE=92209933; PubMed=8458833;
RA   Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT   "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RT   D (CenD), a family A beta-1,4-glucanase.";
RL   J. Bacteriol. 175:1910-1918(1993).
CC   -!- FUNCTION: This enzyme hydrolyzes 1,4-beta-D-glucosidic linkages of
CC       cellulose. Weak activity against carboxymethylcellulose, bacterial
CC       microcrystalline cellulose and barley beta-glucan. Has also weak
CC       endoglucanase activity. Hydrolyzes glucosidic bonds with inversion
CC       of anomeric configuration.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC   -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC       domain.
CC   -!- SIMILARITY: Belongs to cellulase family B (family 6 of glycosyl
CC       hydrolases).
```

```
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11651 / B792;
RX   MEDLINE=95272392; PubMed=7752895;
RA   Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RT   "Candida albicans ALS1 domains related to a Saccharomyces cerevisiae
RT   sexual agglutinin separated by a repeating motif.";
RL   Mol. Microbiol. 15:39-54(1995).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-glycosylated and O-glycosylated (Potential).
CC   -!- SIMILARITY: TO YEAST SAG1.
```

```
DR   EMBL; L25902; AAC41649.2; -.
DR   InterPro; IPR008440; Candida_ALS.
DR   Pfam; PF05792; Candida_ALS; 1.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1260       AGGLUTININ-LIKE PROTEIN 1.
FT   DOMAIN      433    792       10 X 36 AA TANDEM REPEATS.
FT   REPEAT      433    468       1-1.
FT   REPEAT      469    504       1-2.
FT   REPEAT      505    540       1-3.
FT   REPEAT      541    576       1-4.
FT   REPEAT      577    612       1-5.
FT   REPEAT      613    648       1-6.
FT   REPEAT      649    684       1-7.
FT   REPEAT      685    720       1-8.
FT   REPEAT      721    756       1-9.
FT   REPEAT      757    792       1-10.
FT   DOMAIN      983   1152       2 X 26 AA APPROXIMATE REPEATS.
FT   REPEAT      983   1043       2-1.
FT   REPEAT     1092   1152       2-2.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      450    455       POLY-THR.
FT   DOMAIN      491    ...       POLY-THR.
FT   DOMAIN      522    527       POLY-THR.
FT   DOMAIN      558    563       POLY-THR.
FT   DOMAIN      594    599       POLY-THR.
FT   DOMAIN      630    635       POLY-THR.
FT   DOMAIN      666    671       POLY-THR.
FT   DOMAIN      702    707       POLY-THR.
FT   DOMAIN      738    743       POLY-THR.
FT   DOMAIN      774    779       POLY-THR.
FT   DOMAIN      874    877       POLY-SER.
FT   CARBOHYD    471    471       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD    723    723       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD    820    820       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD    886    886       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD    918    918       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD    973    973       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   1045   1045       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   1068   1068       N-LINKED (GLCNAC...)   (POTENTIAL).
SQ   SEQUENCE   1260 AA;  132641 MW;  763D1063A2354C24 CRC64;

Query Match     4.7%;   Score 112.5;   DB 1;   Length 1260;
Best Local Similarity   19.5%;   Pred. No. 3.8;
Matches   68;   Conservative   30;   Mismatches   77;   Indels 173;   Gaps   16;

Qy     5 PNPNPNPGTTTL-SESFENGIPASWKTIDADGDGNN------------------W------ 40
```

```
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   S-layer protein precursor (Surface layer protein).
OS   Lactobacillus brevis.
OC   Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=1580;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=ATCC 8287 / DSM 20556;
RX   MEDLINE=93054358; PubMed=1429463;
RA   Vidgren G., Palva I., Pakkanen R., Lounatmaa K., Palva A.;
RT   "S-layer protein gene of Lactobacillus brevis: cloning by polymerase
RT   chain reaction and determination of the nucleotide sequence.";
RL   J. Bacteriol. 174:7419-7427(1992).
CC   -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC       of proteins which coat the surface of bacteria.
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a
CC       S-layer with tetragonal symmetry.
CC   -!- SIMILARITY: SOME, TO THE S-LAYER PROTEIN OF L.ACIDOPHILUS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z14250; CAA78618.1; -.
DR   PIR; A47023; A47023.
KW   Signal; Glycoprotein; Cell wall; S-layer.
FT   SIGNAL        1     30
FT   CHAIN        31    465       S-LAYER PROTEIN.
SQ   SEQUENCE   465 AA;  48159 MW;  2BE2403392E65A2C CRC64;
```

```
Query Match             4.6%;  Score 111;  DB 1;  Length 465;
Best Local Similarity  19.4%;  Pred. No. 1.3;
Matches   85;  Conservative   45;  Mismatches   187;  Indels  122;  Gaps   16;

Qy     13 TTTLSESFEN-GIPASWKTIDA----DGDGNNWTTTPPPGGTSFAGHNSAICVSSASYIN 67
             ::::.: ::: :::::       :.::
Db     26 TTASAKSYATAGAYSTLKIDAATRNVEATGTNALYTKP--GTV---KGAKVVASKATMAK 80

Qy     68 FEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLE 127
           ::.::  :::::   :::: ::
Db     81 LASSKKSADYFRAYGVKTTNRGSVYYRVTMDGKYRGYVYG-----GKSDTAFAGGI-- 132

Qy    128 EVLTAKTVVTAPEAIRGTRVQGTWYQKTVQ--LPAGTKYVAFRHFGCTDFFWINLDEVEI 185
                                  : .:: :
Db    133 ----KSAETTTKADMPARTTGFYLITDTSKNTLWTAPKYTQYK----------ASKVSL 176

Qy    186 KANGKRADFTETFESSTHGEAPAEWTTIDADGDG-QGWLCLSSGQLDWLTAHGGTNVVAS 244
                  :  : :          ::
Db    177 YGVAKDTKFTVDQAATKTREGSLYYHVTATNGSGISGWIYAGKG---FSTTATGTQVLGG 233

Qy    245 FSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEE 304
             :::::::::::::::::::::::::           :::::::::
Db    234 LSTDK---------SVTATNDNSVKLVVRTTD--------------GTQVGSNTWV--- 266

Qy    305 TNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYV---------AFRHYNCSDLN 356
                                                   :::
Db    267 ----TSTDGTKAGSKVSDKAADQTALEAYINANKPSGYTVTNPNAADAT 311

Qy    357 Y--------ILLDDIQFTMGGSPTPTDYT------------------YTVYRDG 384
            :                                  ::
Db    312 YGNTVYATVSQAATSKVALKVSGTPVTTALTTADANDKVAANDTTANGSSVAGSTVYAAG 371

Qy    385 TKIKEGLTETTFEEDGVAT 403
          ||| :     ||:     :
Db    372 TKLAQLTTDLTGEKGQVVT 390
```

```
DR   EMBL; L25809; AAC36898.1; -.
DR   PIR; S49541; S49541.
DR   HSSP; P07986; 1EXG.
DR   InterPro; IPR001919; Bac_celose-bind.
DR   InterPro; IPR008965; Cellul_bind.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR001524; Glyco_hydro_6.
DR   Pfam; PF00553; CBM_2; 1.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF01341; Glyco_hydro_6; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR00733; GLHYDRLASE6.
DR   ProDom; PD003733; Glyco_hydro_6; 2.
DR   SMART; SM00637; CBD_II; 1.
DR   SMART; SM00060; FN3; 3.
DR   PROSITE; PS00561; CBD_BACTERIAL; 1.
DR   PROSITE; PS00555; GLYCOSYL_HYDROL_F6_1; 1.
DR   PROSITE; PS00556; GLYCOSYL_HYDROL_F6_2; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT   SIGNAL        1     40
FT   CHAIN        41    872       EXOGLUCANASE A.
FT   DOMAIN       41    477       CATALYTIC.
FT   DOMAIN      478    563       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      573    664       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      673    768       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      769    872       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE    188    188       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    410    410       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    140    202       BY SIMILARITY.
FT   DISULFID    374    428       BY SIMILARITY.
FT   DISULFID    770    869       BY SIMILARITY.
SQ   SEQUENCE   872 AA;  89300 MW;  78B3B407F995533B CRC64;
```

```
Query Match             4.7%;  Score 112;  DB 1;  Length 872;
Best Local Similarity  24.0%;  Pred. No. 2.5;
Matches   85;  Conservative   28;  Mismatches  113;  Indels 128;  Gaps   21;

Qy     12 GTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSSASYINFEGP 71
          |||||                  ||:|.  ||:   :::|     ||
Db    586 GTTTTTT-----VPLSWT-----ASTDNAGGSGVAGYE---VLRGTTVV--GT 623

Qy     72 QNPDNYLVTPELSLPGGGTLTFWVCAQD-ANYASEHYAVYASS------------TG 115
          .:  ||| |.:.: ::|                  |||:||::
Db    624 TTATSVTVT--GLTAGTTYSFSVRAKDVAGNTSAASAAVSATTQTGTVVDTTAPSVPTG 680

Qy    116 NDASNFANALLEEVLTAKT------VVTAPEAIRGT--RVQGTWYQKTVQLPAGTKY- 164
                                    ::: ||  ::
Db    681 LTAGTTTTSSVPLTWTASTDNAGGSGVAGYEVFNGTTRVATVTSTSTVT-GLAADTAYS 739

Qy    165 --------------VAFRHFGCTDFFWINLDEVEIKANGKRADFTETFESS 201
                         :  ||         :: |||
Db    740 FTVKAKDVAGNVSAASAAVSARTQAATSGGCT-------VKYSASSWNTGFTGTVEVK 790

Qy    202 THGEAPAEWTTID--ADGD--GQGWLCLSSGQLDWLTAHGGTNVVA-SFSWNGMALNPD 255
           |  | |  |||    ::        |||:||  : :        |
Db    791 NNGTAALNGWTLGFSFADGQKVSQGW-----SAEW--SQSGTAVTAKNAPWNGIL---- 838

Qy    256 NYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGI 309
              :: |||  :      ||    :||
Db    839 -----AAGS------SVSIGFNGTH------NGTN----TAFTAFTLNGV 867
```

```
RESULT 14
SLAP_LACBR
ID   SLAP_LACBR     STANDARD;      PRT;    465 AA.
AC   Q05044;
DT   01-JUN-1994 (Rel. 29, Created)
```

```
RESULT 15
YD94_METJA
ID   YD94_METJA     STANDARD;     PRT;    987 AA.
AC   Q58789;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein MJ1394.
GN   MJ1394.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: TO M.JANNASCHII MJ1393 AND A.FULGIDUS AF2028.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U67579; AAB99404.1; -.
DR   PIR; A64474; A64474.
DR   TIGR; MJ1394;.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     12     32       POTENTIAL.
FT   TRANSMEM    958    978       POTENTIAL.
SQ   SEQUENCE   987 AA;  112360 MW;  D1B628FFB28CA86D CRC64;

Query Match             4.6%;
Best Local Similarity  19.8%;
Matches   81;  Conservative   69;  Mismatches  135;  Indels  124;  Gaps   24;
```

Score 110; DB 1; Length 987;
Pred. No. 4.1;

```
Qy      6 NENPNPGTTTLSESFENGI------PASW--KTIDADG-DGNNWTTTPPPGGTSF-AG  53
            : |::|                 | : |     |||||| :                ||
Db    429 NNDPNGGYKLLPKEIGRDVVISGWVYRPSNWGGGPIDRIGLEDENF-----DGYSFEVN 482

Qy     54 HNSAICVSSASYINFEGPQNPDNLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASS 113
            ||:::  |        |:|:: : ||     ||| :
Db    483 HYS------NYISIDRRTNGNPTEISPEV------YWNPPEDEWYYFE-LKIIYSNG 525

Qy    114 TGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCT 173
               : : |                          ::  ||     :
Db    526 TITFSTYYQNGSL-----AATVST------------IDNT-YTKFDRVVIHGGYVYY------ 564

Qy    174 DFFWINLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWL 233
            | |     |      | : :| || | ::
Db    565 -----VDDLEV--NSKNFDF------YGDKNWKYLEITSANSSEGTAVLFDG--DYF 606

Qy    234 TAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKV----KYYAVND------GFPGD 282
               || :   |  | |   ||:||   ||      :  |      | :
Db    607 KDYNTSNLNAINWTNITLNWSN------DSATLVFNVLGNYSYSERDNILAKYGFA------ 657

Qy    283 HYAVMISKTGTNAGDFTVVFEETPNGINKGGARPGLSTEADGAKPQSVWIERTVDLPAGT 342
               : : |||      | : ||||     :|||
```

```
Db    658 --KILFNYNGTNT------N*TSIKGVVASGS-YSISTDHGTTGEINIWIEN------ 699
            : | ::                  : |  ||| :||:|||| :  ||:|:|:|
Qy    343 KYVAFRH------YNCSDLNYILLDDIQFTMGGSPTFDYTYTVYRDGT 385
            : |                     ||  ::  ||:||       :||||||
Db    700 --VTFKNDAKSYSFNLTNLNIWAVNKSAYELYWNP----FNKSIWIDGS 742
```

Search completed: May 18, 2004, 11:43:37
Job time : 10.2235 secs

```
                Copyright   GenCore version 5.1.6
                           (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:     May 18, 2004, 11:36:20 ; Search time 32.094 Seconds
                                      (without alignments)
                                      4315.838 Million cell updates/sec

Title:      US-08-570-311-22
Perfect score: 2388
Sequence:   1 GTPNPNPNPNPGTTLSESF.........PKVCVNVTINPTQPNPVQNL 439

Scoring table: BLOSUM62
            Gapop 10.0 , Gapext 0.5

Searched:   1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:     1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  SPTREMBL_25:*
             1: sp_archea:*
             2: sp_bacteria:*
             3: sp_fungi:*
             4: sp_human:*
             5: sp_invertebrate:*
             6: sp_mammal:*
             7: sp_mhc:*
             8: sp_organelle:*
             9: sp_phage:*
            10: sp_plant:*
            11: sp_rodent:*
            12: sp_virus:*
            13: sp_vertebrate:*
            14: sp_unclassified:*
            15: sp_rvirus:*
            16: sp_bacteriap:*
            17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	2074	86.9	1723	2	P72197	P72197 porphyromon
2	2068	86.6	1723	2	P72194	P72194 porphyromon
3	2060	86.3	1358	2	P96967	P96967 porphyromon
4	2058	86.2	1687	2	Q9R9B7	Q9R9B7 porphyromon
5	2058	86.2	1704	2	Q51816	Q51816 porphyromon
6	2056.5	86.1	1223	2	Q92NB5	Q92NB5 porphyromon
7	2048	85.8	1706	2	Q51839	Q51839 porphyromon
8	2037	85.3	1706	2	Q51838	Q51838 porphyromon
9	2030.5	85.0	1097	2	P72196	P72196 porphyromon
10	2030.5	85.0	1732	2	007442	007442 porphyromon
11	2025.5	84.8	1732	2	Q51817	Q51817 porphyromon
12	2011.5	84.2	1732	2	052050	052050 porphyromon
13	691.5	29.0	925	2	Q9F4J0	Q9F4J0 porphyromon
14	571.5	23.9	312	2	Q9KIB3	Q9KIB3 porphyromon
15	199	8.3	293	2	Q9XBU9	Q9XBU9 porphyromon
16	136.5	5.7	1341	16	Q8UAU1	Q8UAU1 agrobacteri
17	134.5	5.6	680	2	Q52644	Q52644 ruminococcu
18	134	5.6	2656	5	Q9GNU3	Q9gnu3 paracentrot
19	132.5	5.5	555	16	Q826W1	Q826w1 streptomyce
20	132.5	5.5	1256	16	Q8NX96	Q8nx96 staphylococ
21	132	5.5	1742	16	Q8P377	Q82yw8 xanthomonas
22	131.5	5.5	1744	16	Q82YW8	Q7wbn0 enterococcu
23	131.5	5.5	2215	16	Q7WBN0	Q99v41 bordetella
24	129.5	5.4	1248	16	Q99V41	Q7wtc6 staphylococ
25	129.5	5.4	1255	2	Q7WTC6	Q8e9g6 staphylococ
26	129.5	5.4	2768	16	Q8E9G6	Q7wn54 shewanella
27	129.5	5.4	3346	16	Q7WN54	Q9rzs7 bordetella
28	129	5.4	691	16	Q9RZS7	Q7u7j7 deinococcus
29	128.5	5.4	2230	16	Q7U7J7	Q858b6 synechococc
30	128	5.4	756	9	Q858B6	Q89c73 enterobacte
31	126	5.3	2734	16	Q89C73	Q8pvi0 bradyrhizob
32	124	5.2	1608	17	Q8PVI0	Q8tpz1 methanosarc
33	124	5.2	3988	17	Q8TPZ1	Q8x4h5 methanosarc
34	124	5.2	5188	16	Q8X4H5	Q8x2t1 escherichia
35	124	5.2	5291	16	Q8X2T1	Q7uwz8 escherichia
36	124	5.2	7716	16	Q7UWZ8	085151 rhodopirell
37	123.5	5.2	1095	2	085151	Q8y366 photorhabdu
38	123.5	5.2	1672	16	Q8Y366	Q03658 ralstonia s
39	123.5	5.2	1684	2	Q03658	Q8h753 unidentifie
40	123	5.2	452	10	Q8H753	Q60029 phytophthor
41	123	5.2	596	2	Q60029	Q84dc8 thermomonos
42	123	5.2	1119	2	Q84DC8	Q8eka6 mycoplasma
43	122	5.1	2522	16	Q8EKA6	Q8thc9 shewanella
44	122	5.1	2951	17	Q8THC9	Q9cfa2 methanosarc
45	121	5.1	1649	16	Q9CFA2	lactococcus

ALIGNMENTS

```
RESULT 1
P72197
P72197
ID   P72197      PRELIMINARY;    PRT;   1723 AA.
AC   P72197;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA   Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT   "Molecular cloning and characterization of Porphyromonas gingivalis
RT   Lys-gingipain.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U54691; AAA99810.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE 1723 AA; 186831 MW; 4508A7E50197CEBD CRC64;

Query Match          86.9%;   Score 2074;   DB 2;   Length 1723;
Best Local Similarity  89.1%;   Pred. No. 2.9e-129;
```

SQ   SEQUENCE   1723 AA;   187261 MW;   562896 3D251493EB CRC64;

Query Match
Best Local Similarity   86.6%;   Score 2068;   DB 2;   Length 1723;
                        88.9%;   Pred. No. 7.2e-129;
Matches 391;   Conservative 12;   Mismatches 31;   Indels 6;   Gaps 4;

```
Qy     1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV   60
         ||||||||||||||||||||||||||||||||||| |  ||       |::| ||
Db   969 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 1025

Qy    61 SSASY-INFEGFQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDAS  119
         .:::  :|:  :|:.||| :||.:.|.||| :|||||||||||||||||||||||||||
Db  1026 YSESFGLGGIGVLITPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 1085

Qy   120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN  179
         || .|||||.::|| |   |||||  ||||| ||||| |||||||||||||. ||:|:|:
Db  1086 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 1143

Qy   180 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGT  239
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1144 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGT 1203

Qy   240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  299
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1204 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 1263

Qy   300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL  359
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1264 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 1323

Qy   360 LDDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  419
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1324 LDDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 1383

Qy   420 PKVCVNVTINPTQFNPVQNL  439
         ||.|||||||||||||||.|
Db  1384 PKKCVNVTINPTQFNPVKNL 1403
```

RESULT 3
P96967
ID   P96967   PRELIMINARY;   PRT;   1358 AA.
AC   P96967;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin.
GN   HAGD.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Lepine G., Whitlock J., Wojciechowski L., Progulske-Fox A.;
RT   "Cloning, sequencing and characterization of hagD, a member of the
RT   HArep multigene family in Porphyromonas gingivalis.";
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U68468; AAB49691.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.

Matches 392;   Conservative 11;   Mismatches 31;   Indels 6;   Gaps 4;

```
Qy     1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV   60
         ||||||||||||||||||||||||||||||||||| |  ||       |::| ||
Db   969 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 1025

Qy    61 SSASY-INFEGFQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDAS  119
         .:::  :|:  :|:.||| :||.:.|.||| :|||||||||||||||||||||||||||
Db  1026 YSESFGLGGIGVLITPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 1085

Qy   120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN  179
         || .|||||.::|| |   |||||  ||||| ||||| |||||||||||||. ||:|:|:
Db  1086 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 1143

Qy   180 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGT  239
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1144 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGT 1203

Qy   240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  299
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1204 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 1263

Qy   300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL  359
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1264 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 1323

Qy   360 LDDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  419
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1324 LDDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 1383

Qy   420 PKVCVNVTINPTQFNPVQNL  439
         ||.|||||||||||||||.|
Db  1384 PKKCVNVTINPTQFNPVKNL 1403
```

RESULT 2
P72194
ID   P72194   PRELIMINARY;   PRT;   1723 AA.
AC   P72194;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lys-gingipain.
GN   KGP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=97044756; PubMed=8889827;
RA   Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
RT   "Cloning and sequencing of the gene encoding a novel lysine-specific
RT   cysteine proteinase (Lys-gingipain) in Porphyromonas gingivalis:
RT   structural relationship with the arginine-specific cysteine proteinase
RT   (Arg-gingipain).";
RL   J. Biochem. 120:398-406(1996).
DR   EMBL; D83258; BAA11870.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.

```
DR   SMART; SM00060; FN3; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1358 AA;  147102 MW;  47FCA0B25B06DED8 CRC64;

Query Match             86.3%;  Score 2060;  DB 2;  Length 1358;
Best Local Similarity   88.6%;  Pred. No. 1.8e-128;
Matches  390;  Conservative  12;  Mismatches  32;  Indels  6;  Gaps  4;

Qy      1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICV   60
          ||||||||||||||||||||||||||||||||||||||||||:|     |:|
Db    604 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV  660

Qy     61 SSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDAS  119
          |:  :   |||:||||||||||:||||||||||||||||||||||||||||||||||||
Db    661 YSESFGLGGIGVLTPDNYLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS  720

Qy    120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN  179
          ||:||||||| :|| .:||||||||||||:||||:|||||||||||||||| :||:::
Db    721 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID  778

Qy    180 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGT  239
          |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db    779 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGT  838

Qy    240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    839 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  898

Qy    300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL  359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    899 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL  958

Qy    360 LDDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    959 LDDIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 1018

Qy    420 PKVCVNVTINPTQFNPVQNL  439
          ||||||||||||||||||||
Db   1019 PKVCVNVTINPTQFNPVKNL 1038


RESULT 4
Q9R9B7
ID   Q9R9B7;     PRELIMINARY;     PRT;   1687 AA.
AC   Q9R9B7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemagglutinin/protease.
GN   HAGE.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RA   Han N., Dong H., Progulske-Fox A.;
RT   "Cloning and characterization of hagE from P. gingivalis 381.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF026946; AAD01810.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
```

```
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01164; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE   1687 AA;  183702 MW;  D085B516A399FE70 CRC64;

Query Match             86.2%;  Score 2058;  DB 2;  Length 1687;
Best Local Similarity   88.8%;  Pred. No. 3.2e-128;
Matches  389;  Conservative  12;  Mismatches  31;  Indels  6;  Gaps  4;

Qy      3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICVSS   62
          ||||||||||||||||||||||||||||||||||||||||||:|     |:|
Db    935 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS  991

Qy     63 ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF  121
          |:   |||:||||||||||:||||||||||||||||||||||||||||||||||||
Db    992 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy    122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD  181
          ||||||| :|| .:||||||||||||:||||:|||||||||||||||| :||:::
Db   1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy    182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV  241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1169

Qy    242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1229

Qy    302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1289

Qy    362 DIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 DIQFTMGGSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1349

Qy    422 VCVNVTINPTQFNPVQNL  439
          ||||||||||||||:||
Db   1350 ECVNVTINPTQFNPVKNL 1367


RESULT 5
Q51816
ID   Q51816;     PRELIMINARY;     PRT;   1704 AA.
AC   Q51816;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arg-gingipain-1 proteinase.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95138080; PubMed=7836351;
RA   Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
RA   Travis J., Barr P.J.;
RT   "Molecular cloning and structural characterization of the Arg-
RT   gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a
RT   proteinase-adhesin polyprotein.";
RL   J. Biol. Chem. 270:1007-1010(1995).
DR   EMBL; U15282; AAA69539.1; -.
DR   PIR; A55426; A55426.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
```

```
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   CHAIN       228    719       MATURE 50-KDA CYSTEINE PROTEINASE
FT                                GINGIPAIN.
SQ   SEQUENCE   1704 AA;  1B5436 MW;  6A34B40131C2A676 CRC64;

Query Match             86.2%;
Best Local Similarity   88.8%;   Pred. No. 3.3e-128;
Matches  389;  Conservative  12;  Mismatches  31;  Indels  6;  Gaps  4;

Qy      3  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS    62
           |||||||||||||||||||||||||||||||||||   |         ||||||| ||
Db    952  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS  1008

Qy     63  ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF   121
           |:  ||||||||||||||  ||||||||||||||||||||||| |||||||||||||
Db   1009  ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDASNF  1068

Qy    122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD   181
           :||||||  ||||  :||||||  |||||||||||||||||||||||||||||  ::
Db   1069  TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD  1126

Qy    182  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV   241
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1127  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV  1186

Qy    242  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV   301
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1187  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  1246

Qy    302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD   361
           ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   1247  FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  1306

Qy    362  DIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK   421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307  DIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  1366

Qy    422  VCVNVTINPTQFNPVQNL   439
           :||||||||||||||:||
Db   1367  ECVNVTINPTQFNPVKNL  1384

RESULT 6
Q9ZNB5
ID   Q9ZNB5      PRELIMINARY;      PRT;   1223 AA.
AC   Q9ZNB5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   130k-HMGD (Fragment).
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE=99143166; PubMed=988746;
RA   Shibata Y., Hayakawa M., Takiguchi H., Shiroza T., Abiko Y.;
RT   "Determination and characterization of the hemagglutinin-associated
RT   short motifs found in Porphyromonas gingivalis multiple gene
RT   products.";
RL   J. Biol. Chem. 274:5012-5020(1999).
DR   EMBL; AB019363; BAA34341.1; -
```

```
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   1223 AA;  131542 MW;  00225CD2BA9F91B3 CRC64;

Query Match             86.1%;
Best Local Similarity   88.7%;   Pred. No. 2.7e-128;   DB 2;   Length 1223;
Matches  391;  Conservative  11;  Mismatches  32;  Indels  7;  Gaps  5;

Qy      1  GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC    59
           |||||||||||||  |||||||||||||||||||||||||   |         :||||:|
Db    460  GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC   516

Qy     60  VSSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDA   118
           |:|:|  :|||||||||||||  ||||||||||||||||||||||||| |||||||||||
Db    517  VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVVASSTGNDA   576

Qy    119  SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI   178
           |||:||||||  ||||  :||||||  |||||||||||||||||||||||||||  ::||
Db    577  SNFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI   634

Qy    179  NLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG   238
           :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    635  DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG   694

Qy    239  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF   298
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    695  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF   754

Qy    299  TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI   358
           |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    755  TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI   814

Qy    359  LLDDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV   418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    815  LLDDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV   874

Qy    419  SPKVCVNVTINPTQFNPVQNL   439
           |||||||||||||||||||||
Db    875  SPKVCVNVTINPTQFNPVQNL   895

RESULT 7
Q51839
ID   Q51839      PRELIMINARY;      PRT;   1706 AA.
AC   Q51839; Q51840;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Arginine-specific thiol protease precursor.
GN   PRTR.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=95160709; PubMed=785299;
RA   Kirszbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N.,
RA   Reynolds E.C.;
RT   "Complete nucleotide sequence of a gene prtR of Porphyromonas
```

```
RT   gingivalis W50 encoding a 132 kDa protein that contains an arginine-
RT   specific thiol endopeptidase domain and a haemagglutinin domain.";
RL   Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96311339; PubMed=8713096;
RA   Slakeski N., Cleal S.M., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtR that encodes
RT   an arginine-specific thiol proteinase and multiple adhesins.";
RL   Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Reynolds E.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Slakeski N.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L26341; AAC18876.1; -.
DR   HSSP; P95493; 1CVR.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN       228   1706       ARGININE-SPECIFIC THIOL PROTEASE.
SQ   SEQUENCE   1706 AA;  185626 MW;  E8BDF07C9813B844 CRC64;

Query Match           85.8%;  Score 2048;  DB 2;  Length 1706;
Best Local Similarity 88.1%;  Pred. No. 1.5e-127;
Matches 386;  Conservative 14;  Mismatches 32;  Indels 6;  Gaps 4;

Qy      3  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICVSS   62
           ||||||||||||||||||||||||||||||||:||   ||:|| ||||||||||||||
Db    954  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy     63  ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF  121
           ||: :  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011  ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

Qy    122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD  181
           ||||||||||||||||| ::||: |||:||||||:||:||||||||||||:|: || |
Db   1071  TNALLEETTTAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy    182  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV  241
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1129  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1188

Qy    242  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVV   301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1189  VSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVV  1248

Qy    302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1249  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy    362  DIQFTMGGSPTPTDYTVTVYRDGTKIKBGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1309  DIQFTMGGSPTPTDYTVTVYRDGTKIKBGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368
```

```
Qy    422  VCVNVTINPTQFNPVQNL  439
           ||||||||:| ||||||:||
Db   1369  KCVNVTVNSTQFNPVKNL 1386

RESULT 8
Q51838
ID   Q51838           PRELIMINARY;       PRT;   1706 AA.
AC   Q51838;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Protease precursor.
GN   PRPR1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=96071894; PubMed=7591131;
RA   Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.;
RT   "Characterization, genetic analysis, and expression of a protease
RT   antigen (PrpRI) of Porphyromonas gingivalis W50.";
RL   Infect. Immun. 63:4744-4754(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.;
RT   "The prpR1 and the prR2 arginine-specific protease genes of
RT   Porphyromonas gingivalis W50 produce five biochemically distinct
RT   enzymes.";
RL   Mol. Microbiol. 23:0-0(1997).
DR   EMBL; X82680; CAA57997.1; -.
DR   HSSP; P95493; 1CVR.
DR   MEROPS; C25.001; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN       228    719       ALPHA-PROTEASE.
FT   CHAIN       720   1262       BETA-ADHESIN.
SQ   SEQUENCE   1706 AA;  185705 MW;  0E56DCD87FDA8CDD CRC64;

Query Match           85.3%;  Score 2037;  DB 2;  Length 1706;
Best Local Similarity 87.9%;  Pred. No. 8.2e-127;
Matches 385;  Conservative 13;  Mismatches 34;  Indels 6;  Gaps 4;

Qy      3  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICVSS   62
           ||||||||||||||||||||||||||||||||:||   ||:|| ||||||||||||||
Db    954  PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy     63  ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF  121
           ||: :  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011  ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

Qy    122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD  181
           ||||||||||||||||| ::||: |||:||||||:||:||||||||||||:|: || |
Db   1071  TNALLEETTTAKG-VRSPEAIRG-RIQSTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLL 1128
```

```
Qy    182  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV  241
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1129  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV  1188

Qy    242  VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV   301
             :||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1189  VSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV   1248

Qy    302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1249  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  1308

Qy    362  DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  421
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1309  DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  1368

Qy    422  VCVNVTINPTQFNPVQNL  439
               |||||:|||||:||:|
Db   1369  KCVNVTVNSTQFNPVKNL  1386


RESULT 9
P72196
ID   P72196       PRELIMINARY;      PRT;   1097 AA.
AC   P72196;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   TonB-linked adhesin precursor.
GN   TLA.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9386416; PubMed=9244265;
RA   Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA   Curtis M.A.;
RT   "The tla gene of Porphyromonas gingivalis W50: a homologue of the
RT   arginine-specific protease precursor (PrpRI) which shares sequence
RT   similarity to TonB-linked receptors.";
RL   J. Bacteriol. 179:4788-4788(1997).
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR   EMBL; Y07618; CAA68897.1; -.
DR   GO; GO:0019867; C:outer membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Membrane; Outer membrane; Receptor; Signal; TonB box.
FT   SIGNAL        1     53       POTENTIAL.
SQ   SEQUENCE   1097 AA;  118731 MW;  73BBA337B421F8B9 CRC64;

Query Match             85.0%;  Score 2030.5;  DB 2;  Length 1097;
Best Local Similarity   87.3%;  Pred. No. 1.2e-126;
Matches  385;  Conservative   15;  Mismatches   34;  Indels    7;  Gaps    5;

Qy      1  GTPNPNPNPNGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAIC   59
               |||||||||| | ||||||||||||||||||||||||||:|| || |||||||||||
Db    334  GTPNPNPNPNPNGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC  390

Qy     60  VSSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDA  118
               | ::|  |||   ||||||||||:||||| |||||||||||||||||||||||||||||
```

```
Db    391  VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA  450

Qy    119  SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI  178
               ||| ||||||:|||| :|  : :|||||| || |:||||:||||||||||||||  :|
Db    451  SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI  508

Qy    179  NLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG  238
               :||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    509  DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG  568

Qy    239  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF   298
             :|||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    569  SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF   628

Qy    299  TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI  358
               ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    629  TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI  688

Qy    359  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV  418
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    689  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV  748

Qy    419  SPKVCVNVTINPTQFNPVQNL  439
               |||:|||||:|| :||||:||
Db    749  SPKKCVNVTVNSTQFNPVQNL  769


RESULT 10
O07442
ID   O07442       PRELIMINARY;      PRT;   1732 AA.
AC   O07442;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine-specific cysteine proteinase.
GN   PRTK.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RX   MEDLINE=99935907; PubMed=10219167;
RA   Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RT   "Characterization of a Porphyromonas gingivalis gene prtK that encodes
RT   a lysine-specific cysteine proteinase and three sequence-related
RT   adhesins.";
RL   Oral Microbiol. Immunol. 14:92-97(1999).
DR   EMBL; U75386; AAB60809.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01164; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE   1732 AA;  187914 MW;  45D5B9137739I703 CRC64;

Query Match             85.0%;  Score 2030.5;  DB 2;  Length 1732;
Best Local Similarity   87.3%;  Pred. No. 2.3e-126;
Matches  385;  Conservative   15;  Mismatches   34;  Indels   34;  Gaps    5;

Qy      1  GTPNPNPNPNGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAIC   59
               |||||||||| | |||||||||||||||||||||||||:|| ||  |||||||||||
Db    969  GTPNPNPNPNPNGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC  1025
```

```
Qy    60  VSSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDA  118
              |: :    | ||||:| |: |::|||| : : |:|:|:||||||||||||||||||
Db  1026  VYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy   119  SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI  178
              :||| |||||::| |: : ||||||| :|   ||||||||||||||||:|  |:| |
Db  1086  SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy   179  NLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG  238
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1144  DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy   239  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF  298
              :|:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1204  SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy   299  TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI  358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1264  TVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy   359  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV  418
              ||||||||||||||:||:| ||||||||||||||||||||||||||||||||||||||||
Db  1324  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy   419  SPKVCVNVTINPTQFNPVQNL  439
              |  |:|:|| || ||||||||
Db  1384  SPKKCVDVTVNSTQFNPVQNL 1404
```

```
RESULT 11
Q51817
ID   Q51817           PRELIMINARY;        PRT;    1732 AA.
AC   Q51817;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Porphypain.
GN   PRTP.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W12;
RX   MEDLINE=96213011; PubMed=8631659;
RA   Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA   Proguiske-Fox A., Lantz M.S.;
RT   "Analysis of the prtP gene encoding porphypain, a cysteine proteinase
RT   of Porphyromonas gingivalis.";
RL   J. Bacteriol. 178:2734-2741(1996).
DR   EMBL; U42210; AAB06565.1; -.
DR   PIR; T30836; T30836.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ   SEQUENCE 1732 AA; 187875 MW; 654271DBEF7BCAE4 CRC64;

Query Match           84.8%;  Score 2025.5; DB 2; Length 1732;
Best Local Similarity   87.1%;  Pred. No. 4.9e-126;
Matches  384; Conservative  16; Mismatches  34; Indels  7; Gaps   5;
```

```
Qy    1  GTPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAIC   59
            ||||||||||| |||:|||||||||||||||||||||||     :||:|:|| |
Db  969  GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy   60  VSSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDA  118
            |: :    | ||||:| |: |::||||  : : |:|:|:||||||||||||||||||
Db 1026  VYSESFGLGGIGVLTPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy  119  SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI  178
            :||| |||||::| |: : ||||||| :|   ||||||||||||||||:|  |:| |
Db 1086  SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy  179  NLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG  238
            :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1144  DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy  239  TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF  298
            :|:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1204  SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy  299  TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI  358
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1264  TVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy  359  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV  418
            ||||||||||||||:||:| ||||||||||||||||||||||||||||||||||||||||
Db 1324  LLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy  419  SPKVCVNVTINPTQFNPVQNL  439
            |  |:|:|| || ||||||||
Db 1384  SPKKCVDVTVNSTQFNPVQNL 1404
```

```
RESULT 12
O52050
ID   O52050           PRELIMINARY;        PRT;    1732 AA.
AC   O52050;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lysine specific cysteine protease.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W83;
RX   MEDLINE=98298016; PubMed=9652563;
RA   Lewis J.P., Macrina F.L.;
RT   "IS195, an insertion sequence-like element associated with protease
RT   genes in Porphyromonas gingivalis.";
RL   Infect. Immun. 66:3035-3042(1998).
DR   EMBL; AF017059; AAC26523.1; -.
DR   MEROPS; C25.002; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR001769; Peptidase_C25.
DR   InterPro; IPR005536; Peptidase_C25_C.
DR   Pfam; PF01364; Peptidase_C25; 1.
DR   Pfam; PF03785; Peptidase_C25_C; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW   Protease.
SQ   SEQUENCE 1732 AA; 187931 MW; B237463D5CB5EA5 CRC64;

Query Match           84.2%;  Score 2011.5; DB 2; Length 1732;
Best Local Similarity   87.1%;  Pred. No. 4.9e-126;
Matches  384; Conservative  16; Mismatches  34; Indels  7; Gaps   5;
```

Best Local Similarity 86.6%; Pred. No. 4.2e-125;
Matches 382; Conservative 16; Mismatches 36; Indels 7; Gaps 5;

```
Qy      1 GTPNPNPNPNGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
            ||||||||||||  ||||||||||||||||||||||||||||||      |||||||:
Db    969' GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy     60 VSSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
            |:||: ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1026 VYSESFGLGGIGVLITPDNYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy    119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
            |||:||||||||   :|||||||||||||||||||||||||||||||||||:||
Db   1086 SNFTNALLEEITTAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy    179 NLDEVEIKANGKRADFTETFESSTHGEAPARWTTIDADGDGQGWLCLSSGQLDWLTAHGG 238
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy    239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF 298
            ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy    299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRRHYNCSDLNYI 1323

Qy    359 LLDDIQFTMGGSPTPFDYTYVTVYRDGTKIKEGLTETTEBEDGVATGNHEYCVEVKYTAGV 418
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 LLDDIQFTMGGSPTPFDYTYVTVYRDGTNIKEGLTETTFEEYGVATGNHEYCVEVAKYTAGV 1383

Qy    419 SPKVCVNVTINPTQFNPVQNL 439
            |||||||||||||||||||||
Db   1384 SPKKCVDVTVNSTQPNPVQNL 1404
```

RESULT 13
Q9F4J0

ID   Q9F4J0   PRELIMINARY;   PRT;   925 AA.
AC   Q9F4J0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative outer membrane protein PG57.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA   Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT   "Identification of vaccine candidates from genomic analysis of
RT   Porphyromonas gingivalis.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY007522; AAG42228.1; -.
DR   InterPro; IPR003961; FN_III.
DR   SMART; SM00060; FN3; 1.
SQ   SEQUENCE 925 AA; 103632 MW; 5F2198D6914DAE8 CRC64;

Query Match   29.0%; Score 691.5; DB 2; Length 925;
Best Local Similarity 30.0%; Pred. No. 1.3e-37;
Matches 175; Conservative 63; Mismatches 138; Indels 207; Gaps 20;

```
Qy     16 LSESFENG-IPASWKTIDADGDGNNWTTTPPPGG-----TSFAGHNSAICVSSASYINF 68
```

```
Db    311 LYENFENGPVPNGWLVIDADGDGFSW------GHYLNAYDAPFPGHNGGHCSLSASYVPG 363
            |:||:||::| |||||:|||||||:|:         ::| ||:|| |||::||||||||:
Qy     69 EGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEE 128
            : || ||::||:|:::| ::||:|::::             ::|||||||||:::|||||||
Db    364 IGPVTPDNYLITPKVE--GAKRVKIWVSTQDANWAAEHYAVMASTGTAVGDFV-ILFEE 420

Qy    129 VLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEV--- 183
            :|||        :|:||:|||::|: : ||  ||||||||||||||||||::||:
Db    421 TMTAKPT-----------GAWYERTINLPEGTKYIAWRHYNCTDIYFLKIDDITVFGT 467

Qy    184 --------------------------------------------------EIKANG--- 189
                                                               |:|:|:
Db    468 PASEPPVTDFVVSLIENNKGRLKWNYPNGYEPDKTDDKDPLQLAGYNIYANGSLLVHIQ 527

Qy    190 --------------KRADFTET------------------------------FESSTHGE 205
                          |:|||:||                              |||||||:
Db    528 DPTVLEYIDETYSSRDDQVEVEYCVTAVYNDNIESQSVCDKLIYDSQSDIILYEGFEAGS 587

Qy    206 APAEWTTIDADGDGQGWLCLSSGQLDW-LTAHGGTNVVASFSWNGM-ALNPDNYLISKD 262
            :|||||||||||||||||||||||||  |:    :::| : ||||   |||||||:|:
Db    588 IPEGWLLIDADGDNVNW------------DYYPWTMYGHDSEKCIASPSYLPMIGVLTPDNYLVTPR 642

Qy    263 VTGATKVKYYAVNDG-FPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTE 321
             :|:::|:||:|  : ::                        |:|:||||||:||:|||:
Db    643 LEGAKLVKYWVSAQDAVYSAEHYAVNVSTTGTAVEDFVLFEET------------MTAK 690

Qy    322 ADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDI------------------ 363
            |:|:         :||:|:|||||||||||:|||::||||||
Db    691 ANGA---WYERTTLPAGTKYIAWRHYDCTDMFFLLLDDITVRSTETVPEPVTDFVV 745

Qy    364 --------------QFTMGGSPTPTD------YTYTVYRDGT-----------KIKEGLTETTF 396
                          |||||||:|:|:       :|:|||||           |:|:||:|:|:
Db    746 SLIENNKGRLKWNYPNGYEPDKTDDKKPLQLTGYNIYANGSLLVHIQDPPTVLEYIDETYS 805

Qy    397 EEDGVATGNHEYCVEVKYTAGV-SPKVCVNVTINPTQFNPVQN 438
            : ||    :|||:|||:::||:   :  |:  ||::|:|||:|:
Db    806 SRDGQV--EMEYCVTAVYNDNIESQSVCDKLNYTITSLDNIQS 846
```

RESULT 14
Q9KIB3

ID   Q9KIB3   PRELIMINARY;   PRT;   312 AA.
AC   Q9KIB3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical outer membrane protein PG27.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C.;
RT   "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF237555; AAF81413.1; -.
SQ   SEQUENCE 312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;

Query Match   23.9%; Score 571.5; DB 2; Length 312;
Best Local Similarity 33.3%; Pred. No. 3e-30;
Matches 138; Conservative 23; Mismatches 71; Indels 183; Gaps 8;

```
Qy     16 LSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSSASYINFEGPQNPD 75
```

```
                        ||||||:|||| ||||||||||| ||         ||:| | :||||||||   ||
Db          4 LSESFESGIPAIWKTIDADGDGYNWMHL-----TNFTG-QSGLCVSSASYIGGVGALTPD  57

                        |||:|||||| |   :|||||      |  :|||||:|  |  | ||||||
Qy         76 NYLVTPELSLPGGG--TLTFWVCAQDANYASHYAVYASSTGNDASNFANALLEEVLTAK 133

                        |||:||||:|   : :|||        ||||:||:|||::|||::|:| | |||||
Db         58 NYLITPELKLPTDALVEIIYWVCTQDLTAPSEHYAVYSSSTGNNAADFVNLLYEETLTAK 117

Qy        134 TVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFPWINLLDEVEIKANGKRAD 193
                        : :|| ||||  |||: ||:| ||:||||| || ||:|||:||
Db        118 R-IQSPELIRGNRTQGVWYQRKVVLPNDTKYVAFRHFNSTDNFWLNLDEVSIL------- 169

Qy        194 FTETFSSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALN 253
                        |
Db        170 ------------------------------------------------------------ 169

Qy        254 PDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVFEETPNGINKGG 313
                                                                       |:
Db        170 ------------------------------------------------YTP-------- 172

Qy        314 ARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTP 373
                                                                 ||  |
Db        173 ---------------------------------LP------RRAPC---------PHP 182

Qy        374 TDYTYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTI 428
                        ||:|:||| |   ||||||||   : |:||   ::|| ||||:|||:| : :
Db        183 GGYTYSVFRDGGQKIASGLASGLASALAVIDTDVPYGTQDYCVQVNYLQGDSYKVCKNIVV 237
```

```
RESULT 15
Q9XBU9
ID   Q9XBU9        PRELIMINARY;        PRT;     293 AA.
AC   Q9XBU9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Immunoreactive 32 kDa antigen PG25.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Porphyromonadaceae; Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "P. gingivalis polypeptides and nucleic acids.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF155351; AAD38410.1; -.
SQ   SEQUENCE  293 AA;  32272 MW;  CC03EAC241F7F6F1 CRC64;
```

```
Query Match        8.3%;   Score 199;   DB 2;   Length 293;
Best Local Similarity   60.9%;   Pred. No. 1.7e-05;
Matches   39;  Conservative   5;  Mismatches   18;  Indels   2;  Gaps   1;

Qy        376 YTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNVTINPTQFNP 435
                        |||||||||||| ||||||| | :|| | ||||||||||||||||||:|::
Db         77 YTYTVYRDGVKIASGLTEPTFLDEDVPAGEHTYCVEVQYQGGVSDKVCVDVEVK--DFKP 134

Qy        436 VQNL 439
                        |  |
Db        135 VINL 138
```

```
Search completed: May 18, 2004, 11:46:41
Job time : 34.094 secs
```

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:38:40 ; Search time 13.5635 Seconds
(without alignments)
1670.936 Million cell updates/sec

Title: US-08-570-311-22
Perfect score: 2388
Sequence: 1 GTPNPNPNPGTTTLSESF.........PKVCVNVTINPTQFNPVQNL 439

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2388	100.0	439	2	US-08-570-311-22	Sequence 22, Appl
2	2388	100.0	2628	2	US-08-570-311-14	Sequence 18, Appl
3	2351	98.5	456	2	US-08-570-311-18	Sequence 18, Appl
4	2351	98.5	456	2	US-08-570-311-20	Sequence 20, Appl
5	2321	97.2	450	2	US-08-570-311-16	Sequence 16, Appl
6	2058	86.2	1477	4	US-09-482-500A-1	Sequence 1, Appli
7	2058	86.2	1687	2	US-08-570-311-29	Sequence 29, Appl
8	2058	86.2	1704	2	US-08-336-308A-10	Sequence 6, Appl
9	2058	86.2	1704	3	US-08-822-324-6	Sequence 6, Appli
10	2058	86.2	1704	3	US-09-490-931-10	Sequence 10, Appl
11	2055	86.1	1087	2	US-08-570-311-8	Sequence 8, Appli
12	2055	86.1	1087	2	US-08-353-485-8	Sequence 8, Appli
13	2055	86.1	1358	2	US-08-570-311-27	Sequence 27, Appl
14	2048	85.8	1706	4	US-09-066-330-11	Sequence 11, Appl
15	2030.5	85.0	1732	4	US-09-066-330-11	Sequence 11, Appl
16	2025.5	84.8	1732	2	US-08-570-311-10	Sequence 10, Appl
17	2025.5	84.8	1732	2	US-08-353-485-10	Sequence 10, Appl
18	458.5	19.2	497	2	US-08-570-311-2	Sequence 2, Appli
19	458.5	19.2	497	2	US-08-353-485-2	Sequence 2, Appli
20	294.5	12.3	942	4	US-08-141-324-14	Sequence 14, Appl
21	294.5	12.3	942	1	US-08-541-902-14	Sequence 14, Appl
22	242	10.1	49	3	US-08-822-324-18	Sequence 18, Appl
23	195	8.2	46	2	US-08-822-324-9	Sequence 9, Appli
24	131.5	5.5	1638	4	US-09-071-035-258	Sequence 258, App
25	131.5	5.5	1638	4	US-09-071-035-262	Sequence 262, App
26	131.5	5.5	1638	4	US-09-071-035-266	Sequence 266, App
27	131.5	5.5	1747	4	US-09-134-000C-5999	Sequence 5999, Ap
28	129	5.4	24	4	US-09-066-330-6	Sequence 6, Appli
29	128	5.4	25	3	US-08-822-324-19	Sequence 19, Appl
30	125.5	5.3	669	4	US-09-071-035-264	Sequence 264, App
31	125	5.2	25	3	US-08-822-324-14	Sequence 14, Appl
32	124	5.2	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
33	123.5	5.2	1095	4	US-08-851-567B-34	Sequence 34, Appl
34	120	5.0	1833	4	US-08-621-944A-4	Sequence 4, Appli
35	120	5.0	1833	4	US-08-945-567D-4	Sequence 4, Appli
36	120	5.0	1992	4	US-08-621-944A-3	Sequence 3, Appli
37	120	5.0	1992	3	US-08-945-567D-3	Sequence 3, Appli
38	118.5	5.0	2123	3	US-08-968-685A-10	Sequence 10, Appl
39	118	4.9	24	3	US-08-822-324-20	Sequence 20, Appl
40	118	4.9	2353	3	US-09-377-155-33	Sequence 33, Appl
41	118	4.9	2353	3	US-08-913-942-4	Sequence 4, Appli
42	118	4.9	2353	4	US-09-669-974-33	Sequence 33, Appl
43	118	4.9	2353	4	US-09-797-862-33	Sequence 33, Appl
44	118	4.9	2411	4	US-09-268-347-36	Sequence 36, Appl
45	117.5	4.9	331	1	US-08-535-230A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-570-311-22
; Sequence 22, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:

```
;      LENGTH: 439 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-570-311-22

Query Match            100.0%;  Score 2388; DB 2;  Length 439;
Best Local Similarity  100.0%;  Pred. No. 2.7e-204;
Matches  439; Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV 60

Qy     61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy    121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHPGCTDFFWINL 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHPGCTDFFWINL 180

Qy    181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTN 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTN 240

Qy    241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

Qy    301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

Qy    361 DDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 DDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy    421 KVCVNVTINPTQFNPVQNL 439
            |||||||||||||||||||
Db    421 KVCVNVTINPTQFNPVQNL 439

RESULT 2
US-08-570-311-14
; Sequence 14, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;      APPLICANT: Progulske-Fox, Ann
;      APPLICANT: Tumwasorn, Somying
;      APPLICANT: Lepine, Guylaine
;      APPLICANT: Han, Naiming
;      APPLICANT: Lantz, Marilyn
;      APPLICANT: Patti, Joseph
;      TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;      TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;      NUMBER OF SEQUENCES: 29
;      CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Ted W. Whitlock
;      STREET: 2421 N.W. 41st Street, Suite A-1
;      CITY: Gainesville
;      STATE: FL
;      COUNTRY: USA
;      ZIP: 32606
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: Patentin Release #1.0, Version #1.25
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/570,311
;      FILING DATE:
;      CLASSIFICATION: 424
```

```
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/353,485
;      FILING DATE: 09-DEC-1994
;      CLASSIFICATION: 424
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 07/647,119
;      FILING DATE: 25-JAN-1991
;      CLASSIFICATION: 424
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 07/241,640
;      FILING DATE: 08-SEP-1988
;      ATTORNEY/AGENT INFORMATION:
;      NAME: Whitlock, Ted W.
;      REGISTRATION NUMBER: 36,965
;      REFERENCE/DOCKET NUMBER: UF15.C3
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (904) 375-8100
;      TELEFAX: (904) 372-5800
;  INFORMATION FOR SEQ ID NO: 14:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 2628 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-570-311-14

Query Match            100.0%;  Score 2388; DB 2;  Length 2628;
Best Local Similarity  100.0%;  Pred. No. 4.2e-203;
Matches  439; Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1862 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV 1921

Qy     61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1922 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1981

Qy    121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHPGCTDFFWINL 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1982 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHPGCTDFFWINL 2041

Qy    181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTN 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2042 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTN 2101

Qy    241 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2102 VVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 2161

Qy    301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2162 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 2221

Qy    361 DDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2222 DDIQFTMGGSPTPTDYTYTVVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 2281

Qy    421 KVCVNVTINPTQFNPVQNL 439
            |||||||||||||||||||
Db   2282 KVCVNVTINPTQFNPVQNL 2300

RESULT 3
US-08-570-311-18
; Sequence 18, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;      APPLICANT: Progulske-Fox, Ann
;      APPLICANT: Tumwasorn, Somying
;      APPLICANT: Lepine, Guylaine
;      APPLICANT: Han, Naiming
;      APPLICANT: Lantz, Marilyn
```

```
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 18:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 456 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-18

Query Match            98.5%;  Score 2351;  DB 2;  Length 456;
Best Local Similarity  98.2%;  Pred. No. 5.7e-201;
Matches 431;  Conservative  3;  Mismatches 3;  Indels 5;  Gaps  0;

Qy   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60

Qy  61 SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
         |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db  61 SSASYINFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

Qy 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

Qy 181 DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN 240
         | ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db 181 DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLGWLTAHGGTN 240

Qy 241 VVASFSWNGMALNPDNYLLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGNAGDFTV 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 VVASFSWNGMALNPDNYLLISKDVTGATKVKYYYAVNDGFPGDHYAVMLSKTGNAGDFTV 300

Qy 301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLIPAGTKYVAFRHYNCSDLNYILL 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLIPAGTKYVAFRHYNCSDLNYILL 360
```

```
Qy 361 DDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 DDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP 420

Qy 421 KVCVNVTINPTQFNPVQNL 439
         |||| :: |||||||
Db 421 KECVNVTVDPVQFNPVQNL 439

RESULT 4
US-08-570-311-20
; Sequence 20, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: .US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/647,119
;   FILING DATE: 25-JAN-1991
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 456 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-570-311-20

Query Match            98.5%;  Score 2351;  DB 2;  Length 456;
Best Local Similarity  98.2%;  Pred. No. 5.7e-201;
Matches 431;  Conservative  3;  Mismatches 3;  Indels 5;  Gaps  0;

Qy   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60
```

```
                                              REGISTRATION NUMBER: 36,965
                                         REFERENCE/DOCKET NUMBER: UF15.C3
                                      TELECOMMUNICATION INFORMATION:
                                         TELEPHONE: (904) 375-8100
                                         TELEFAX:   (904) 372-5800
                                 INFORMATION FOR SEQ ID NO: 16:
                                    SEQUENCE CHARACTERISTICS:
                                      LENGTH: 450 amino acids
                                      TYPE: amino acid
                                      TOPOLOGY: linear
                                    MOLECULE TYPE: protein
                                 US-08-570-311-16

Query Match             97.2%;  Score 2321;  DB 2;  Length 450;
Best Local Similarity   98.4%;  Pred. No. 2.6e-198;
Matches 426;  Conservative   3;  Mismatches   4;  Indels   0;  Gaps   0;

Qy    7  PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSPAGHNSAICVSSASYI   66
         |||||||||||||||||||||||||||||||||||||||||| |||||||| |||||||
Db    1  PNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICASSASYI   60

Qy   67  NFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL  126
         |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   61  NFEGPQNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL  120

Qy  127  EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIK  186
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db  121  EEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK  180

Qy  187  ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS  246
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  ANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFS  240

Qy  247  WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP  306
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  WNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETP  300

Qy  307  NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT  366
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  NGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT  360

Qy  367  MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKVCVNV  426
         |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  361  MGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNV  420

Qy  427  TINPTQFNPVQNL  439
         | :: ||||||||
Db  421  TVDPVQFNPVQNL  433


RESULT 6
US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
;   APPLICANT: Travis, James
;   APPLICANT: Imamura, Takahisa
;   APPLICANT: Potempa, Jan
;   TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
;   FILE REFERENCE: 235.00160101
;   CURRENT APPLICATION NUMBER: US/09/482,500A
;   CURRENT FILING DATE: 2001-12-17
;   PRIOR APPLICATION NUMBER: US 60/115,869
;   PRIOR FILING DATE: 1999-01-13
;   NUMBER OF SEQ ID NOS: 4
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 1477
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-09-482-500A-1

Query Match             86.2%;  Score 2058;  DB 4;  Length 1477;
```

```
Qy   61  SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  SSASYINFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN  120

Qy  121  FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL  180

Qy  181  DEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN  240
         | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  DDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTN  240

Qy  241  VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  VVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTV  300

Qy  301  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  VFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL  360

Qy  361  DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  DDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSP  420

Qy  421  KVCVNVTINPTQFNPVQNL  439
         | ||||| :: |||||||
Db  421  KECVNVTVDPVQFNPVQNL  439


RESULT 5
US-08-570-311-16
; Sequence 16, Application US/08570311
; Patent No. 5824791
GENERAL INFORMATION:
  APPLICANT: Progulske-Fox, Ann
  APPLICANT: Tumwasorn, Somying
  APPLICANT: Lepine, Guylaine
  APPLICANT: Han, Naiming
  APPLICANT: Lantz, Marilyn
  APPLICANT: Patti, Joseph
  TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
  TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
  NUMBER OF SEQUENCES: 29
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Ted W. Whitlock
    STREET: 2421 N.W. 41st Street, Suite A-1
    CITY: Gainesville
    STATE: FL
    COUNTRY: USA
    ZIP: 32606
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/570,311
    FILING DATE:
    CLASSIFICATION: 424
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/353,485
    FILING DATE: 09-DEC-1994
    CLASSIFICATION: 424
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/647,119
    FILING DATE: 25-JAN-1991
    CLASSIFICATION: 424
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/241,640
    FILING DATE: 08-SEP-1988
  ATTORNEY/AGENT INFORMATION:
    NAME: Whitlock, Ted W.
```

Best Local Similarity   88.8%;   Pred. No. 4.3e-174;
Matches 389; Conservative 12; Mismatches 31; Indels   6; Gaps   4;

Qy   3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICVSS   62
            ||||||||||||||||||||||||||||||||||||: : ||||||||||||
Db   725 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 781

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
            | :| ||||||||||||||||||||||||||||||||||||||||||||||
Db   782 ESFGLGGIGVLTPDNYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNF 841

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
            ||||| :|||| :||||| |:|||||||| ||||||||||||:|||: |:|
Db   842 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 899

Qy   182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 241
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   900 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 959

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   960 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1019

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1020 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1079

Qy   362 DIQFTMGGSPTPDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1080 DIQFTMGGSPTPDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1139

Qy   422 VCVNVTINPTQFNPVQNL 439
            ||||||||||||:|
Db   1140 ECVNVTINPTQFNPVKNL 1157

Query Match   86.2%;   Score 2058; DB 2;   Length 1687;
Best Local Similarity 88.8%;   Pred. No. 5.2e-174;
Matches 389; Conservative 12; Mismatches 31; Indels   6; Gaps   4;

Qy   3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICVSS   62
            ||||||||||||||||||||||||||||||||||||: : ||||||||||||
Db   935 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 991

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
            | :| ||||||||||||||||||||||||||||||||||||||||||||||
Db   992 ESFGLGGIGVLTPDNYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNF 1051

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
            ||||| :|||| :||||| |:|||||||| ||||||||||||:|||: |:|
Db   1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1109

Qy   182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 241
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   :110 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1169

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   :170 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1229

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1289

Qy   362 DIQFTMGGSPTPDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 DIQFTMGGSPTPDYTYTYVRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1349

Qy   422 VCVNVTINPTQFNPVQNL 439
            ||||||||||||:|
Db   1350 ECVNVTINPTQFNPVKNL 1367

GENERAL INFORMATION:
APPLICANT: Potempa, Jan S.
APPLICANT: Travis, James
APPLICANT: Genco, Caroline A.
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,324
FILING DATE: 21-MAR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/013,945
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 103-95 WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 488-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-822-324-6

Query Match   86.2%;   Score 2058; DB 3; Length 1704;
Best Local Similarity 88.8%;   Pred. No. 5.3e-174;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

```
Qy    3  PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS   62
            |||||||||||||||||||:||  ||||
Db  952  PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy   63  ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF  121
            |:  ||||||||||||||||||||||||
Db 1009  ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy  122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD  181
            :|  ||||||| :|  :|||
Db 1069  TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy  182  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV  241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1127  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy  242  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1187  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy  302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
            ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db 1247  FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy  362  DIQFTMGGSPTPDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  421
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1307  DIQFTMGGSPTPDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366
```

COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,308A
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/119,361
FILING DATE: 10-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,441
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 21-93C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-308A-10

Query Match   86.2%;   Score 2058; DB 3; Length 1704;
Best Local Similarity 88.8%;   Pred. No. 5.3e-174;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

```
Qy    3  PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS   62
            |||||||||||||||||||:||  ||||
Db  952  PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy   63  ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF  121
            |:  ||||||||||||||||||||||||
Db 1009  ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy  122  ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD  181
            :|  ||||||| :|  :|||
Db 1069  TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy  182  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV  241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1127  EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy  242  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV  301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1187  VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1246

Qy  302  FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD  361
            ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db 1247  FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy  362  DIQFTMGGSPTPDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK  421
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1307  DIQFTMGGSPTPDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy  422  VCVNVTINPTQFNPVQNL  439
            ||||||||||:|
Db 1367  ECVNVTINPTQFNPVKNL 1384
```

RESULT 9
US-08-822-324-6
; Sequence 6, Application US/08822324
; Patent No. 6129917

Qy   422 VCVNVTINPTQFNPVQNL 439
            ||||||||||||||||:||
Db   1367 ECVNVTINPTQFNPVKNL 1384

RESULT 10
US-09-490-931-10
; Sequence 10, Application US/09490931
; Patent No. 6274718
; GENERAL INFORMATION:
;   APPLICANT: Travis, James
;   APPLICANT: Potempa, Jan S.
;   APPLICANT: Barr, Philip J.
;   APPLICANT: Pavloff, Nadine
;   TITLE OF INVENTION: Porphyromonas gingivalis
;   TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
;   STREET: 5370 Manhattan Circle, Suite 201
;   CITY: Boulder
;   STATE: Colorado
;   COUNTRY: US
;   ZIP: 80303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/490,931
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/336,308
;   FILING DATE:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/265,441
;   FILING DATE: 24-JUN-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ferber, Donna M.
;   REGISTRATION NUMBER: 33,878
;   REFERENCE/DOCKET NUMBER: 21-93C
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (303) 499-8080
;   TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1704 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-490-931-10

Query Match   86.2%;   Score 2058;   DB 3;   Length 1704;
Best Local Similarity 88.8%; Pred. No. 5.3e-174;
Matches   389; Conservative 12; Mismatches 31; Indels   6; Gaps   4;

Qy   3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS 62
          ||||||||||||||||||||||||||||||||||||||| || :| |||||
Db   952 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1008

Qy   63 ASY-INFBGPQNPDNYLVTPELSLPGGGTLITFWVCAQDANYASEHYAVYASSTGNDASNF 121
          ||| |||||||||||||||:||| ||||||||||||||||||||||||||||||||||
Db   1009 ESFGLGGIGVLTPDNYLITPALDLPNGGKLITFWVCAQDANYASEHYAVYASSTGNDASNF 1068

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          ||: ||| ||||||| |||| :|||||||||||:|||||||||||||||| |:|:||
Db   1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1126

Qy   182 EVEIKANGKRADFTETEFSSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   1127 EVEIKANGKRADFTETEFSSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNV 1186

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVV 301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1187 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGNAGDFTVV 1246

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1247 FEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306

Qy   362 DIQFTMGGSPTPTDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307 DIQFTMGGSPTPTDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1366

Qy   422 VCVNVTINPTQFNPVQNL 439
            ||||||||||||||||:||
Db   1367 ECVNVTINPTQFNPVKNL 1384

RESULT 11
US-08-570-311-8
; Sequence 8, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
;   APPLICANT: Progulske-Fox, Ann
;   APPLICANT: Tumwasorn, Somying
;   APPLICANT: Lepine, Guylaine
;   APPLICANT: Han, Naiming
;   APPLICANT: Lantz, Marilyn
;   APPLICANT: Patti, Joseph
;   TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
;   TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ted W. Whitlock
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/570,311
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,485
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/241,640
;   FILING DATE: 08-SEP-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Whitlock, Ted W.
;   REGISTRATION NUMBER: 36,965
;   REFERENCE/DOCKET NUMBER: UF15.C3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1087 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein

US-08-570-311-8

Query Match   86.1%; Score 2055; DB 2; Length 1087;
Best Local Similarity 88.4%; Pred. No. 4.9e-174;
Matches 389; Conservative 13; Mismatches 32; Indels 6; Gaps 4;

```
Qy    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV   60
          ||||||||||||||||||||||||||||||||||||      :|| ||  |:||
Db  333 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV  389

Qy   61 SSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDAS  119
          ::|:  :|| | | :||:||  :|:|  || ||||||||||||||||||||||||||||
Db  390 YSESFGLGGIGVLTPDNYLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS  449

Qy  120 NFANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN  179
          || ||||| :|||| :|  ||||| ||:|||:|||:||||||||||||||  || |:|
Db  450 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID  507

Qy  180 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGT  239
          ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  508 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGT  567

Qy  240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  568 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  627

Qy  300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL  359
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||:||||
Db  628 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYIL  687

Qy  360 LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  688 LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  747

Qy  420 PKVCVNVTINPTQFNPVQNL  439
          |||||||||||||||||:||
Db  748 PKVCVNVTINPTQFNPVKNL  767
```

RESULT 12
US-08-353-485-8
; Sequence 8, Application US/08353485
; Patent No. 5830710
GENERAL INFORMATION:
APPLICANT: Progulske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1087 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-353-485-8

Query Match   86.1%; Score 2055; DB 2; Length 1087;
Best Local Similarity 88.4%; Pred. No. 4.9e-174;
Matches 389; Conservative 13; Mismatches 32; Indels 6; Gaps 4;

```
Qy    1 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV   60
          ||||||||||||||||||||||||||||||||||||      :|| ||  |:||
Db  333 GTPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV  389

Qy   61 SSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDAS  119
          ::|:  :|| | | :||:||  :|:|  || ||||||||||||||||||||||||||||
Db  390 YSESFGLGGIGVLTPDNYLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS  449

Qy  120 NFANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN  179
          || ||||| :|||| :|  ||||| ||:|||:|||:||||||||||||||  || |:|
Db  450 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID  507

Qy  180 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGT  239
          ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  508 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQDWLCLSSGQLDWLTAHGGT  567

Qy  240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  568 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT  627

Qy  300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL  359
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||:||||
Db  628 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYIL  687

Qy  360 LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  688 LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS  747

Qy  420 PKVCVNVTINPTQFNPVQNL  439
          |||||||||||||||||:||
Db  748 PKVCVNVTINPTQFNPVKNL  767
```

RESULT 13
US-08-570-311-27
; Sequence 27, Application US/08570311
; Patent No. 5824791
GENERAL INFORMATION:
APPLICANT: Progulske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville

```
          STATE: FL
          COUNTRY: USA
          ZIP: 32606
    COMPUTER READABLE FORM:
          MEDIUM TYPE: Floppy disk
          COMPUTER: IBM PC compatible
          OPERATING SYSTEM: PC-DOS/MS-DOS
          SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
          APPLICATION NUMBER: US/08/570,311
          FILING DATE:
          CLASSIFICATION: 424
    PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US 08/353,485
          FILING DATE: 09-DEC-1994
          CLASSIFICATION: 424
    PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US 07/647,119
          FILING DATE: 25-JAN-1991
          CLASSIFICATION: 424
    PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US 07/241,640
          FILING DATE: 08-SEP-1988
    ATTORNEY/AGENT INFORMATION:
          NAME: Whitlock, Ted W.
          REGISTRATION NUMBER: 36,965
          REFERENCE/DOCKET NUMBER: UF15.C3
    TELECOMMUNICATION INFORMATION:
          TELEPHONE: (904) 375-8100
          TELEFAX: (904) 372-5800
    INFORMATION FOR SEQ ID NO: 27:
    SEQUENCE CHARACTERISTICS:
          LENGTH: 1358 amino acids
          TYPE: amino acid
          TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-570-311-27

Query Match    86.1%;    Score 2055;  DB 2;   Length 1358;
Best Local Similarity  88.4%;  Pred. No. 6.9e-174;
Matches 389;  Conservative 13;  Mismatches 32;  Indels 6;  Gaps 4;

Qy    1 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV 60
          ||||||||||||||||||||||||||||||||||||||| :|||       ||:|
Db  604 GTPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 660

Qy   61 SSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
          :::   |||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  661 YSESFGLGGIGVLTPDNYLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 720

Qy  120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWIN 179
          || :|||||||||||| :|| |||||||||||||:||||||||||||||||||| ||||
Db  721 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 778

Qy  180 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGT 239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  779 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGT 838

Qy  240 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  839 NVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFT 898

Qy  300 VVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYIL 359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  899 VVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLDYIL 958

Qy  360 LDDIQFTMGGSPTPTDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  959 LDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVS 1018

Qy  420 PKVCVNVTINPTQFNPVQNL 439
          |||||||||||||:|||:|
Db 1019 PKVCVNVTINPTQFNPVKNL 1038
```

```
RESULT 14
US-09-066-330-10
;  Sequence 10, Application US/09066330A
;  Patent No. 6511666
;  GENERAL INFORMATION:
;  APPLICANT: Reynolds, Eric C.
;  APPLICANT: Bhogal, Peter S.
;  APPLICANT: Slakeski, Nada
;  TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;  FILE REFERENCE: Reynolds
;  CURRENT APPLICATION NUMBER: US/09/066,330A
;  CURRENT FILING DATE: 1998-09-15
;  EARLIER APPLICATION NUMBER: PN 6275
;  EARLIER FILING DATE: 1995-10-30
;  EARLIER APPLICATION NUMBER: PCT/AU96/00673
;  EARLIER FILING DATE: 1996-10-30
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 10
;  LENGTH: 1706
;  TYPE: PRT
;  ORGANISM: Porphyromonas gingivalis
US-09-066-330-10

Query Match    85.8%;    Score 2048;  DB 4;   Length 1706;
Best Local Similarity  88.1%;  Pred. No. 4.1e-173;
Matches 386;  Conservative 14;  Mismatches 32;  Indels 6;  Gaps 4;

Qy    3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICVSS 62
          |||||||||||||||||||||||||||||||||||||||| :|||       ||:|||
Db  954 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy   63 ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          : :   |||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db 1011 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

Qy  122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
          :|||||||||||| :|| |||||||||||||:||||||||||||||||||| ||||:|:|
Db 1071 TNALLEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy  182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNV 241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1129 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNV 1188

Qy  242 VASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
          :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy  302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1249 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy  362 DIQFTMGGSPTPTDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1309 DIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy  422 VCVNVTINPTQFNPVQNL 439
          |||||||||||:||:|
Db 1369 KCVNVTVNSTQFNPVKNL 1386
```

```
RESULT 15
US-09-066-330-11
;  Sequence 11, Application US/09066330A
;  Patent No. 6511666
;  GENERAL INFORMATION:
;  APPLICANT: Reynolds, Eric C.
;  APPLICANT: Bhogal, Peter S.
;  APPLICANT: Slakeski, Nada
;  TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;  FILE REFERENCE: Reynolds
```

```
;  CURRENT APPLICATION NUMBER: US/09/066,330A
;  CURRENT FILING DATE: 1998-09-15
;  EARLIER APPLICATION NUMBER: PN 6275
;  EARLIER FILING DATE: 1995-10-30
;  EARLIER APPLICATION NUMBER: PCT/AU96/00673
;  EARLIER FILING DATE: 1996-10-30
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;  LENGTH: 1732
;  TYPE: PRT
;  ORGANISM: Porphyromonas gingivalis
US-09-066-330-11

Query Match             85.0%;   Score 2030.5;   DB 4;  Length 1732;
Best Local Similarity   87.3%;   Pred. No. 1.5e-171;
Matches    385; Conservative  15; Mismatches  34; Indels   7; Gaps   5;

Qy      1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC  59
          ||||||||||||| ||||||||||||||||||||||||:   |:  ||:|
Db    969 GTPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy     60 VSSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
          ||||:| :  |||:|||||||:|||||||||||||||||||||||||||||||||||||
Db   1026 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 1085

Qy    119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
          ||| |||||||||:|| :|||| |: :||||||||||||||||||||||||| |:
Db   1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy    179 NLDEVEIKANGKRADFTETESSSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 238
          :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1144 DLDEVEIKANGKRADFTETFSSSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGG 1203

Qy    239 TNVVASFSWNGMALNPDNYLIISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 298
          :||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 SNVVSSFSWNGMALNPDNYLIISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy    299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy    359 LLDDIQFTMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 LLDDIQFTMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy    419 SPKVCVNVTINPTQFNPVQNL 439
          |||:||||||:| ||||||||
Db   1384 SPKKCVNVTVNSTQFNPVQNL 1404
```

Search completed: May 18, 2004, 11:49:17
Job time : 14.5635 secs

```
            GenCore version 5.1.6
         Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May 18, 2004, 11:42:50 ; Search time 34.0043 Seconds
                                       (without alignments)
                                       3592.387 Million cell updates/sec

Title:        US-08-570-311-22
Perfect score: 2388
Sequence:     1 GTPNPNPNPNPGTTTLSESF.........PKVCVNVTINPTQFNPVQNL 439

Scoring table: BLOSUM62
              Gapop 10.0 , Gapext 0.5

Searched:     1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: .    1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications AA:*
```

Published Applications list (1–18):

```
 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2048	85.8	1706	14	US-10-229-066-10	Sequence 10, Appl
2	2030.5	85.0	1732	14	US-10-229-066-11	Sequence 11, Appl
3	727	30.4	419	15	US-10-174-695-3	Sequence 3, Appli
4	711.5	29.8	419	15	US-10-174-695-5	Sequence 5, Appli
5	303	12.7	231	15	US-10-174-695-6	Sequence 4, Appli
6	233.5	9.8	196	15	US-10-174-695-4	Sequence 18, Appl
7	153	6.4	29	15	US-10-387-977-18	Sequence 21, Appl
8	153	6.4	29	15	US-10-387-977-21	Sequence 23, Appl
9	153	6.4	29	15	US-10-387-977-23	Sequence 24, Appl
10	153	6.4	29	15	US-10-387-977-24	Sequence 26, Appl
11	153	6.4	29	15	US-10-387-977-26	Sequence 17, Appl
12	147	6.2	29	15	US-10-387-977-17	Sequence 25, Appl
13	145	6.1	29	15	US-10-387-977-25	Sequence 81, Appl
14	137	5.7	25	15	US-10-387-977-81	Sequence 20, Appl
15	136	5.7	29	15	US-10-387-977-20	Sequence 20, Appl
16	133	5.6	16	12	US-10-282-122A-55467	Sequence 55467, A
17	132.5	5.5	14	14	US-10-156-761-14592	Sequence 14592, A
18	132	5.5	15	15	US-10-387-977-83	Sequence 83, Appl
19	132	5.5	15	15	US-10-387-977-84	Sequence 84, Appl
20	132	5.5	15	15	US-10-387-977-85	Sequence 85, Appl
21	131.5	5.5	14	15	US-10-311-879-15	Sequence 15, Appl
22	131.5	5.5	12	15	US-10-206-576-258	Sequence 258, App
23	131.5	5.5	12	15	US-10-206-576-262	Sequence 262, App
24	131.5	5.5	12	15	US-10-206-576-266	Sequence 266, App
25	131.5	5.5	12	15	US-10-282-122A-56997	Sequence 56997, A
26	131	5.5	12	15	US-10-282-122A-47453	Sequence 47453, A
27	130	5.4	15	15	US-10-387-977-77	Sequence 77, Appl
28	130	5.4	25	15	US-10-387-977-80	Sequence 80, Appl
29	129	5.4	24	14	US-10-229-066-6	Sequence 6, Appli
30	129	5.4	691	15	US-10-369-493-23473	Sequence 23473, A
31	125.5	5.3	669	12	US-10-206-576-264	Sequence 264, App
32	125	5.2	25	15	US-10-387-977-76	Sequence 76, Appl
33	125	5.2	25	15	US-10-387-977-79	Sequence 79, Appl
34	123.5	5.2	1095	15	US-10-242-056-34	Sequence 34, Appl
35	123.5	5.2	1095	15	US-10-262-794A-34	Sequence 34, Appl
36	121	5.1	1649	15	US-10-369-493-18460	Sequence 18460, A
37	120	5.0	1833	14	US-10-175-282-4	Sequence 4, Appli
38	120	5.0	1833	14	US-10-175-275-4	Sequence 4, Appli
39	120	5.0	1992	14	US-10-175-282-3	Sequence 3, Appli
40	120	5.0	1992	14	US-10-175-275-3	Sequence 3, Appli
41	119	5.0	1333	12	US-10-282-122A-55546	Sequence 55546, A
42	118.5	5.0	2122	9	US-09-813-214A-9	Sequence 9, Appli
43	118	4.9	2353	9	US-09-797-862-33	Sequence 33, Appl
44	117	4.9	22	14	US-10-229-066-4	Sequence 4, Appli
45	117	4.9	29	15	US-10-387-977-19	Sequence 19, Appl

ALIGNMENTS

```
RESULT 1
US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US20031015637A1
; GENERAL INFORMATION:
;  APPLICANT: Reynolds, Eric C.
;  APPLICANT: Bhogal, Peter S.
;  APPLICANT: Slakeski, Nada
;  TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;  FILE REFERENCE: Reynolds
;  CURRENT APPLICATION NUMBER: US/10/229,066
;  CURRENT FILING DATE: 2002-08-28
;  PRIOR APPLICATION NUMBER: US/09/066,330
;  PRIOR FILING DATE: 1998-09-15
;  PRIOR APPLICATION NUMBER: PN 6275
;  PRIOR FILING DATE: 1995-10-30
;  PRIOR APPLICATION NUMBER: PCT/AU96/00673
;  PRIOR FILING DATE: 1996-10-30
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;  LENGTH: 1706
;  TYPE: PRT
;  ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match          85.8%;   Score 2048;  DB 14;  Length 1706;
Best Local Similarity  88.1%;  Pred. No. 1e-177;
Matches  386;  Conservative  14;  Mismatches  32;  Indels  6;  Gaps  4;

Qy     3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTPPPGGTSFAGHNSAICVSS 62
           ||||||||||||||||||||||||||||||||||||||||||      ||||||||
Db   954 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVYS 1010

Qy    63 ASY-INFEGPQNPDNYLVTPELSLPCGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
          ||:  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1011 ESFGLGGIGVLITPDNYLITPALDLPNGGKITFWVCAQDANYASEHYAVYASSTGNDASNF 1070
```

```
Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
         |||||    |||| |    |  :|:|||||:|||:||||||||||||||  ||:|:|:||
Db  1071 TNALLEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 1128

Qy   182 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNV 241
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1129 EVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGGTNV 1188

Qy   242 VASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 301
         | :|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTVV 1248

Qy   302 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1249 FEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1308

Qy   362 DIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1309 DIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368

Qy   422 VCVNVTINPTQFNPVQNL 439
         :||||| |:||:||||:|
Db  1369 KCVNVTVNSTQFNPVKNL 1386
```

RESULT 2
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US2003015763TA1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric C.
;   APPLICANT: Bhogal, Peter S.
;   APPLICANT: Slakeski, Nada
;   TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
;   FILE REFERENCE: Reynolds
;   CURRENT APPLICATION NUMBER: US/10/229,066
;   CURRENT FILING DATE: 2002-08-28
;   PRIOR APPLICATION NUMBER: US/09/066,330
;   PRIOR FILING DATE: 1998-09-15
;   PRIOR APPLICATION NUMBER: PN 6275
;   PRIOR FILING DATE: 1995-10-30
;   PRIOR APPLICATION NUMBER: PCT/AU96/00673
;   PRIOR FILING DATE: 1996-10-30
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1732
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

Query Match   85.0%; Score 2030.5; DB 14; Length 1732;
Best Local Similarity   87.3%; Pred. No. 4.3e-176;
Matches 385; Conservative 15; Mismatches 34; Indels 7; Gaps 5;

```
Qy     1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAIC 59
         |||||||||||||  ||||||||||||||||||||||   ||:||:||| ||:
Db   969 GTPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 1025

Qy    60 VSSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
         |::|:  |||||||||||||||||||:|| || ||||||||||||||||||||||||||
Db  1026 VTSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANVASEHYAVYASSTGNDA 1085

Qy   119 SNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWI 178
         |||:||||| |||| |    |  :|:|||||:|||:||||||||||||||  ||:|:|:
Db  1086 SNFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143

Qy   179 NLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGG 238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1144 DLDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQWLCLSSGQLDWLTAHGG 1203

Qy   239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGFPGDHYAVMISKTGTNAGDF 298
          :||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1204 SNVVSSFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDF 1263

Qy   299 TVVFEETPNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1264 TVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323

Qy   359 LLDDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 418
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1324 LLDDIQFTMGGSPTPTDYTTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGV 1383

Qy   419 SPKVCVNVTINPTQFNPVQNL 439
         |||:||||| |:||:||||||
Db  1384 SPKKCVNVTVNSTQFNPVQNL 1404
```

RESULT 3
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US2003023022A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: Slakeski, Nada
;   APPLICANT: Chen, Chao Guang
;   APPLICANT: Barr, Ian George
;   TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
;   FILE REFERENCE: 529282000700
;   CURRENT APPLICATION NUMBER: US/10/174,695
;   CURRENT FILING DATE: 2002-06-18
;   PRIOR APPLICATION NUMBER: PCT/AU00/01588
;   PRIOR FILING DATE: 2000-12-21
;   PRIOR APPLICATION NUMBER: AU PQ 4859
;   PRIOR FILING DATE: 1999-12-24
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

Query Match   30.4%; Score 727; DB 15; Length 419;
Best Local Similarity   75.8%; Pred. No. 1.1e-57;
Matches 144; Conservative 10; Mismatches 30; Indels 6; Gaps 4;

```
Qy     3 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSS 62
         |||||||||||||||||||||||||||||||||   ||:||:||| ||:
Db   235 PNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCVVS 291

Qy    63 ASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
         |::  |||||||||||||||||:|| || ||||||||||||||||||||||||||||||
Db   292 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 351

Qy   122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLD 181
         |||| |    |  :|:|||||:|||:||||||||||||||  ||:|:|:||
Db   352 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLD 409

Qy   182 EVEIKANGKR 191
         ||||||||||
Db   410 EVEIKANGKR 419
```

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US2003023022A1
; GENERAL INFORMATION:
;   APPLICANT: Reynolds, Eric Charles
;   APPLICANT: Slakeski, Nada
;   APPLICANT: Chen, Chao Guang
;   APPLICANT: Barr, Ian George
;   TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
;   FILE REFERENCE: 529282000700
;   CURRENT APPLICATION NUMBER: US/10/174,695

```
Db   92 RYDDFFFE--AGKKYTFTMRRAGMGDGTDMEVED-------DSPASYTYVYRDGTKIK 141
          |||| ||          ||| |  |||||  :|
Qy  389 EGLTETTFEEDGVATGNHEYCVEVKTAGVSPXVCVANVTI-NPTQFNPVQNL 439
          |||| |||||||  ||||||||||| |||||| ||  |    :||  | :||||
Db  142 EGLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNL 193
```

RESULT 6
US-10-174-695-4
; Sequence 4, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-4

Query Match   9.8%; Score 233.5; DB 15; Length 196;
Best Local Similarity 75.9%; Pred. No. 5.2e-13;
Matches 44; Conservative 2; Mismatches 9; Indels 3; Gaps 1;

Qy   3 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAICV 60
          |||||||||||||||||||||||||||||||||| |||||
Db   44 PNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGCV 98
```

RESULT 7
US-10-387-977-18
; Sequence 18, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-18

Query Match      6.4%;  Score 153;  DB 15;  Length 29;
Best Local Similarity 100.0%;  Pred. No. 7.3e-07;
Matches  29;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

```
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-5
```

```
Query Match     29.8%;  Score 711.5; DB 15;  Length 419;
Best Local Similarity  74.1%;  Pred. No. 2.8e-56;
Matches 143;  Conservative 11;  Mismatches 32;  Indels  7;  Gaps  5;

Qy    1 GTPNPNPNPNPGT-TTLSESFENGIPASWKTIDADGDGNNWTTTPPGGTSFAGHNSAIC 59
          ||||||||||||  ||||||||||||||||||||||||   ||  |||
Db  232 GTPNPNPNPNPNPGTTLSESFENGIPASWKTIDADGDGHGWKPGNAPG---IAGYNSNGC 288

Qy   60 VSSASY-INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDA 118
          |:  :|   |:
Db  289 VYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDA 348

Qy  119 SNFANALLEEVLTAKTVVTAPRAIRGRVQGTWYQKYTVQLPAGTKYVAFRHFGCTDFFWI 178
          |||:|||||:  |||  |  | ||| ||||| ||| :||  |||||||||||| |||  |
Db  349 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVDLPLPAGTKYVAFRHFQSTDMFYI 406

Qy  179 NLDEVEIKANGKR 191
          :||||||||||||
Db  407 DLDEVEIKANGKR 419
```

RESULT 5
US-10-174-695-6
; Sequence 6, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-6

```
Query Match     12.7%;  Score 303;  DB 15;  Length 231;
Best Local Similarity  37.5%;  Pred. No. 2.9e-19;
Matches  87;  Conservative 30;  Mismatches 69;  Indels 46;  Gaps 11;

Qy  213 IDADGDGQGWLCLSSGQLDWLTAHGGTNVV-ASFSWNGMALNPDNVLISKD--VTGATKV 269
          :|||:      |:     |  :|  :||
Db    3 LDADHNTFGSVIPATGPL--PTGTASSNLYSANFEVL-IPANADPVVTQNIIVTGQGEV 59

Qy  270 KYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEADGAKPQS 329
                           :|
Db   60 V-------IPGGVY-------DYCITNPEPASG------KMWIAGDGGNQPA 91

Qy  330 VWIERTVDLPAGTKYV-AFRHYNCSDINYILLDDIOFTMGGSPTPTDYTYTVYRDGTKIK 388
          :| :||  |||||||        ||     |       ||||||||||||||
```

```
Qy          375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
                    ||||||||||||||||||||||||||||
Db            1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29
```

RESULT 8
US-10-387-977-21
; Sequence 21, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-21

Query Match   6.4%;   Score 153; DB 15; Length 29;
Best Local Similarity   100.0%; Pred. No. 7.3e-07;
Matches   29; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
                    ||||||||||||||||||||||||||||
Db            1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29
```

RESULT 9
US-10-387-977-23
; Sequence 23, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-23

Query Match   6.4%;   Score 153; DB 15; Length 29;
Best Local Similarity   100.0%; Pred. No. 7.3e-07;
Matches   29; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
                    ||||||||||||||||||||||||||||
Db            1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29
```

RESULT 10
US-10-387-977-24
; Sequence 24, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-24

Query Match   6.4%;   Score 153; DB 15; Length 29;
Best Local Similarity   100.0%; Pred. No. 7.3e-07;
Matches   29; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
                    ||||||||||||||||||||||||||||
Db            1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29
```

RESULT 11
US-10-387-977-26
; Sequence 26, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-26

Query Match   6.4%;   Score 153; DB 15; Length 29;
Best Local Similarity   100.0%; Pred. No. 7.3e-07;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy  375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
            |||||||||||||||||||||||||||||
Db    1 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 29
```

RESULT 12
US-10-387-977-17
; Sequence 17, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-17

Query Match 6.2%; Score 147; DB 15; Pred. No. 2.6e-06;
Best Local Similarity 96.6%; Matches 28; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

```
Qy  375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
            |||||||||||||||||||||||||||
Db    1 DYTYTVYRDGTKIKEGLTATTFEEDGVAT 29
```

RESULT 13
US-10-387-977-25
; Sequence 25, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-25

Query Match 6.1%; Score 145; DB 15; Length 29;

Best Local Similarity 96.6%; Pred. No. 3.9e-06;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy  375 DYTYTVYRDGTKIKEGLTETTFEEDGVAT 403
            ||||||||||||||| |||||||||||
Db    1 DYTYTVYRDGTKIKWGLTETTFEEDGVAT 29
```

RESULT 14
US-10-387-977-81
; Sequence 81, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 81
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-81

Query Match 5.7%; Score 137; DB 15; Pred. No. 1.7e-05;
Best Local Similarity 100.0%; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy  404 GNHEYCVEVKYTAGVSPKVCVNVTI 428
            |||||||||||||||||||||||||
Db    1 GNHEYCVEVKYTAGVSPKVCVNVTI 25
```

RESULT 15
US-10-387-977-20
; Sequence 20, Application US/10387977
; Publication No. US2004005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-20

```
Query Match              5.7%;  Score 136;  DB 15;  Length 29;
Best Local Similarity   96.3%;  Pred. No. 2.6e-05;
Matches   26;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   0;

Qy    376 YTYTVYRDGTKIKBGLTETTFEEDGVA 402
          |||||||||||||||| |||||||||
Db      2 YTYTVYRDGTKIKEGLTATTFEEDGVA 28

Search completed: May 18, 2004, 11:52:22
Job time : 35.0043 secs
```